Macroscopic behavior of heterogenous populations with fast random life histories

Alexandre Boumezoued

Macroscopic behavior of heterogenous populations with fast random life histories[1]

Alexandre Boumezoued[2]

December 16, 2015

Abstract

In this paper, we consider the large population limit of an age and characteristic-structured stochastic population model evolving according to individual birth, death and fast characteristics changes during life. Both the large population framework and the fast characteristics changes assumption are motivated by demographic patterns of human populations at the scale of a given country. When rescaling the population process, and under some invariance assumption about the characteristics changes dynamics, the classical deterministic transport-renewal McKendrick-Von Foerster equation appears, that describes the time evolution of the age pyramid driven by equivalent birth and death rates. The proof follows the work of Méléard and Tran (2012) and Gupta et al. (2014) in which analogous mathematical issues are encountered. We further prove that the sequence of processes taking track of the characteristics distribution is not tight even in the presence of age-independent demographic rates. To illustrate the use of the limiting model, a set of computable invariant distributions is given, as well as numerical implementation of equivalent birth and death rates which mimics real demographic data. These results highlight the fact that characteristics changes frequencies are crucial to understand aggregate demographic rates at the macroscopic scale.

Keywords: Population dynamics, mathematical demography, age pyramid, heterogeneity, multi-state models, birth-death processes, point processes, limit theorems, fast life trajectories.

1 Introduction

The study of the dynamics of mortality rates at the national level is still a major issue both in demographics and actuarial science. Whereas national data provides

[1]This work benefited from the financial support of the ANR project *Lolita* (ANR-13-BS01-0011).

[2]Laboratoire de Probabilités et Modèles Aléatoires (LPMA), UMR CNRS 7599
Université Paris 6, 4 Place Jussieu, 75005 Paris, France. Email: alexandre.boumezoued@upmc.fr

estimation of death rates for men and women by age over time, a growing interest concerns the dynamics of mortality rates at a deeper level, looking at other individual characteristics. This is crucial to understand the dynamics of mortality, both for state pension issues and insurance risk assessment. In particular, the statistical estimation of death rates by individual characteristics can been carried out when looking at a sample of individuals who share some characteristics which are quite stable during their life. However in practice, many characteristics of individuals can vary over time. A possible approach is to estimate a "larger" model including death rates but also transitions between characteristics or states. The statistical techniques dedicated to so-called multi-state models address this issue for a class of processes and observation schemes (see e.g. Boumezoued et al. (2015) and references therein). In this field however, the statistical estimation remains very challenging, partly due to the lack of data and the censoring scheme, but also due to the high number of involved parameters. Another approach could be to deal with an approximate model. Indeed, among the characteristics which are known to have a real impact on longevity, many vary very frequently over time: one can think to exposition or not to some contaminant, alternating a dangerous activity with a safe one, the precise income, the health status with alternate periods of illness. Those can also have a huge impact on birth patterns, making the dynamics of the whole population difficult to analyze. As an example, the time evolution of the age pyramid is a crucial quantity of interest for decision making in public pension systems. In this context of fast changing characteristics, it seems difficult to keep track of the population evolution at the microscopic level and one could be interested in the right "approximation" of the dynamics, that is, to replace all birth and death rates by characteristics by an aggregate death rate which depends on the microscopic rates but also on some stable population composition.

In this paper, we consider a stochastic population model in which individuals have an age and characteristics, and can give birth, change their characteristics (event called swap) and/or die. We study the asymptotics of the stochastic individual-based model under both large population and fast swap patterns; in particular characteristics change at the fast time scale whereas aging remain at the slow time scale. Both the large population framework and the frequent characteristics changes appear naturally when focusing on the demographic evolution of a human population at the scale of a given country. In this context, and under some invariance assumption on the swap patterns, the macroscopic behavior is described by a McKendrick-Von Foerster deterministic equation (see McKendrick (1926) and Von Foerster (1959)) in which only age is involved and parameters are averaged over the stable characteristics distribution. Our probabilistic setting is inspired by Fournier and Méléard (2004), Champagnat et al. (2006), Tran (2008), Ferriere and

Tran (2009) and Bensusan et al. (2010–2015). In particular, the birth-death-swap process representation is based on Bensusan et al. (2010–2015) (see also Bensusan (2010)).

In the literature, limit theorems for stochastic processes involving several time scales have been widely studied. As for our framework of interest which concerns measure-valued population processes, one can find several studies involving two time scales, for example related to evolutionary mechanisms with rare or accelerated mutations in characteristics-structured (called trait-structured) population models (see e.g. Fournier and Méléard (2004), Champagnat et al. (2006), Bovier and Wang (2013) and Billiard et al. (2014)), prey-predator models (see e.g. Costa et al. (2015) and Costa (2015)), as well as fast aging structured populations (see e.g. Méléard and Tran (2012) and Gupta et al. (2014)). To our knowledge, no contribution focused on measure-valued age and trait-structured population models with fast changing characteristics and age-dependent birth and death rates. This is the purpose of the present paper to develop such modeling framework, as well as to highlighting its contribution to demographic purposes. This work can be seen as the probabilistic counterpart of aggregation methods for deterministic equations based on time scale separation techniques (see e.g. Auger et al. (2012) for a review). Concerning the rescaling techniques, we are specifically interested in Méléard and Tran (2012) and Gupta et al. (2014) in which analogous mathematical issues are encountered, and whose techniques are used in the present paper. The link with our result will be further detailed in the corresponding section.

The remainder of this paper is organized as follows. In Section 2, the birth-death-swap population process is introduced, as well as the assumptions and our results. Proofs are given in Section 3. Finally, Section 4 details some examples and numerical illustration, and we give some concluding remarks in Section 5.

2 Setting and main results

We are interested in the evolution of a population in which each individual has characteristics $x \in \mathcal{X}$, where $\mathcal{X}$ is a compact set of $\mathbb{R}^d$, and an age $a \in \mathbb{R}_+$. In the population, three kinds of events can occur:

(i) A birth, that is the arrival of an individual with age zero,

(ii) A death, that is the removal of an individual,

(iii) A swap, that is a change of individual's characteristics.

We want to model the fact that each individual changes its characteristics very often compared to the times at which it gives birth and dies. This is motivated by human populations for which one can consider individuals who change their income, health status, or food condition very often during their life. Each individual with characteristics x and age a gives birth at rate $b(x,a)$, dies with rate $d(x,a)$, and changes its characteristics at times given by the swap rate $n.e(x,a)$, where n is the scale parameter which will be grown to infinity. We are interested in the macroscopic behavior of such population evolution, so the scale parameter n is also used as the order of magnitude of the population size.

We consider a reference probability measure $l(\mathrm{d}x)$ on the space $\mathcal{X}$ of characteristics. At birth, mutations occur thanks to a kernel $k_b(x,a,x')l(\mathrm{d}x')$. At a time of swap, the characteristics x of the individual with age a are replaced by new characteristics x' drawn according to the kernel $k_e(x,a,x')l(\mathrm{d}x')$.

2.1 Construction

We construct the population processes as solutions to a thinning problem. This construction for measure-valued birth-death processes can be found in Fournier and Méléard (2004), Champagnat et al. (2006), Tran (2008), Ferriere and Tran (2009), as well as in particular Bensusan et al. (2010–2015) for the detailed construction of a birth-death-swap process. Let $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space satisfying the usual conditions. On this probability space, let $Q(\mathrm{d}s, \mathrm{d}i, \mathrm{d}x', \mathrm{d}\theta)$ be a Poisson point measure on $\mathbb{R}_+ \times \mathbb{N}^* \times \mathcal{X} \times \mathbb{R}_+$ with intensity measure $\mathrm{d}s n(\mathrm{d}i) l(\mathrm{d}x')\mathrm{d}\theta$, where $n(.)$ is the counting measure on $\mathbb{N}^*$, that is $n(A)$ is the number of elements in A. On this probability space, let us also introduce for each $n \in \mathbb{N}^*$ the initial population, represented by a random point measure $Z_0^n(\mathrm{d}x, \mathrm{d}a) = \sum_{i=1}^{N_t^0} \delta_{X^i(Z_0), A^i(Z_0)}(\mathrm{d}x, \mathrm{d}a)$ on the space $\mathcal{X} \times \mathbb{R}_+$ which puts a weight on characteristics and ages of individuals present at time 0. Let $(\mathcal{F}_t)$ be the canonical filtration generated by Z_0 and Q, which will be the reference filtration in this paper. The population at time t indexed by the scale parameter $n \in \mathbb{N}^*$ is denoted $Z_t^n(\mathrm{d}x, \mathrm{d}a) = \sum_{i=1}^{N_t^n} \delta_{X^i(Z_t), A^i(Z_t)}(\mathrm{d}x, \mathrm{d}a)$, where at any time individuals are ordered by age. The virtue of the measure representation is that one can compute a function of the whole population structure using the notation

$$\langle Z_t^n, f \rangle = \int_{\mathcal{X} \times \mathbb{R}_+} f(x,a) Z_t^n(\mathrm{d}x, \mathrm{d}a) = \sum_{i=1}^{N_t^n} f\left(X^i(Z_t), A^i(Z_t)\right).$$

For example, the population size is $N_t^n = \langle Z_t^n, \mathbf{1} \rangle$, whereas the total population birth intensity is $\langle Z_{t-}^n, b \rangle$. The population process is constructed as the solution to a *thinning* problem which can be interpreted as follows: the population at time t is computed as the population Z_0 corrected by all random demographic events which happened between 0 and t, namely birth, death and swap events. More precisely, for each $n \in \mathbb{N}^*$, the measure-valued process $Z_.^n(\mathrm{d}x, \mathrm{d}a)$ is defined as the solution to the following equation:

$$
\begin{aligned}
Z_t^n(\mathrm{d}x,\mathrm{d}a) &= \sum_{j=1}^{\langle Z_0^n,\mathbf{1}\rangle} \delta_{(X^i(Z_0^n),A^i(Z_0^n)+t)}(\mathrm{d}x,\mathrm{d}a)\\
&+ \int_0^t \int_{N^*}\int_{\mathcal{X}}\int_{\mathbb{R}_+} \mathbf{1}_{i\le\langle Z_{s-}^n,\mathbf{1}\rangle}\Big(\mathbf{1}_{0\le\theta<m_1(Z_{s-}^n,i,x')}\delta_{(x',t-s)}(\mathrm{d}x,\mathrm{d}a)\\
&+ \mathbf{1}_{m_1(Z_{s-}^n,i,x')\le\theta<m_2(Z_{s-}^n,i,x')}\Big(\delta_{(x',A^i(Z_{s-}^n)+t-s)}(\mathrm{d}x,\mathrm{d}a)-\delta_{(X^i(Z_{s-}^n),A^i(Z_{s-}^n)+t-s)}(\mathrm{d}x,\mathrm{d}a)\Big)\\
&- \mathbf{1}_{m_2(Z_{s-}^n,i,x')\le\theta<m_3(Z_{s-}^n,i,x')}\delta_{(X^i(Z_{s-}^n),A^i(Z_{s-}^n)+t-s)}(\mathrm{d}x,\mathrm{d}a)\Big)Q(\mathrm{d}s,\mathrm{d}i,\mathrm{d}x',\mathrm{d}\theta),
\end{aligned}
\tag{1}
$$

where $m_1(Z_{s-}^n,i,x') = b(X^i(Z_{s-}^n),A^i(Z_{s-}^n))k_b(X^i(Z_{s-}^n),A^i(Z_{s-}^n),x')$, $m_2(Z_{s-}^n,i,x') = m_1(Z_{s-}^n,i,x') + n.e(X^i(Z_{s-}^n),A^i(Z_{s-}^n))k_e(X^i(Z_{s-}^n),A^i(Z_{s-}^n),x')$, and $m_3(Z_{s-}^n,i,x') = m_2(Z_{s-}^n,i,x') + d(X^i(Z_{s-}^n),A^i(Z_{s-}^n))$.

Existence and uniqueness results are given in Subsection 2.3. Note that by construction the measure-valued process has the Markov property, since the total intensity at time t is fully determined by the population Z_{t-}. In order to study the macroscopic behavior, let us first define the renormalized measure

$$\tilde{Z}_t^n(\mathrm{d}x,\mathrm{d}a) := \frac{1}{n}Z_t^n(\mathrm{d}x,\mathrm{d}a),$$

that is, the population in which each individual has weight $1/n$. The corresponding age pyramid processes of interest are defined below.

Definition 1. *(Age pyramid) The sequence of measure-valued process $(\bar{Z}_t^n(\mathrm{d}a))_{t\ge0}$ defined as the age marginal by: for each $f\in C_b(\mathbb{R}_+)$ (continuous and bounded on $\mathbb{R}_+$), $\int_{\mathbb{R}_+} f(a)\bar{Z}_t^n(\mathrm{d}a) = \int_{\mathcal{X}\times\mathbb{R}_+} f(a)\tilde{Z}_t^n(\mathrm{d}x,\mathrm{d}a)$ (denoted $\langle\bar{Z}_t^n,f\rangle = \langle\tilde{Z}_t^n,f\rangle$) is called sequence of age pyramid processes.*

We are interested in the limit as n grows to infinity, so the frequency of swaps $n.e(x,a)$ increases to infinity whereas the birth and death rates remain the same. The idea is to separate the time scale of demographic events (birth and death), and that of characteristics changes. At the same time, the population will be grown to infinity with n (see Assumption 4 below), so that two effects remain:
(i) As the population grows, the stochastic dynamics averages to a deterministic pattern,
(ii) As swap events occur more and more frequently, under some invariance assumption, some stable distribution of the characteristics is reached by the whole population.
Then, in the limit $n\to+\infty$, the age pyramid process should solve a deterministic equation in which demographic parameters are averaged over the stable distribution of the characteristics.

2.2 Assumptions

We detail here our assumptions and we discuss their use and interpretation.

Assumption 1. *(Bounded intensities and mutation densities) Demographic rates and mutation kernels are continuous and there exists positive constants* $\bar{k}$, $\bar{b}$, $\bar{d}$ *and* $\bar{e}$ *such that for all* $(x, a, x') \in \mathcal{X} \times \mathbb{R}_+ \times \mathcal{X}$, $k_e(x, a, x') \leq \bar{k}$, $k_b(x, a, x') \leq \bar{k}$, $b(x, a) \leq \bar{b}$, $d(x, a) \leq \bar{d}$ *and* $e(x, a) \leq \bar{e}$.

Assumption 2. *(Control of moments) There exists* $\alpha > 0$ *such that*

$$\sup_{n \geq 1} \mathbb{E}\left[\left(\langle \tilde{Z}_0^n, \mathbf{1} \rangle\right)^{1+\alpha}\right] < +\infty.$$

Remark 1. *The proof can be adapted (and simplified) in the case where ages lie in* $[0, \bar{a}]$, *where* $\bar{a}$ *is some limiting age. In this framework, one can check that only the assumption that* $\sup_n \mathbb{E}\left[\langle \tilde{Z}_0^n, \mathbf{1} \rangle\right] < +\infty$ *is required. However, such limit age would constraint the model and is in fact restrictive for many applications. Therefore, our aim is to detail a more general convergence result for an unconstrained age space. This aspect will be discussed in the proof Section 3.*

The crucial assumption about the invariance of the swap pattern is given below. It states that at the individual level, the way characteristics change during life admits an invariant distribution. Moreover, we propose here a general framework in which this invariant pattern may depend on age.

Assumption 3. *(Invariant measure) For each* $a \in \mathbb{R}_+$, *there exists a positive solution* $x \mapsto g(x, a)$ *such that* $\int_{\mathcal{X}} g(x, a) l(\mathrm{d}x) = 1$ *to the Fredholm equation:* $l(\mathrm{d}x)$*-a.e.,*

$$e(x, a) g(x, a) = \int_{\mathcal{X}} g(y, a) e(y, a) k_e(y, a, x) l(\mathrm{d}y). \tag{2}$$

Remark 2. *(Probabilistic interpretation of Assumption 3) We omit age for the discussion here. Consider a Markov process with values in* $\mathcal{X}$, *which jumps from a state* x *to a state* y *with rate* $e(x) k_e(x, y)$. *Its infinitesimal generator is given by* $\mathcal{A}f(x) = e(x) \left(\int_{y \in \mathcal{X}} k_e(x, y) f(y) l(\mathrm{d}y) - f(x)\right)$ *for each continuous* f. *It is easy to see that under Assumption 3, for each continuous* f, $\int_{\mathcal{X}} \mathcal{A}f(x) g(x) l(\mathrm{d}x) = 0$. *If the Markov process is right-continuous, then the measure* $\nu(\mathrm{d}x) = g(x) l(\mathrm{d}x)$ *is its unique invariant measure (up to scaling).*

Remark 3. *Equation (2) is of the form* $\Psi(x) = \int_{y \in \mathcal{X}} \Psi(y) k_e(y, x) l(\mathrm{d}y)$ *(for each age) and is called homogenous Fredholm equation of the second kind. For the study of the solutions to such equations, we refer to Zemyan (2012). In Section 4, we provide examples of solutions both for a mixture kernel and a model with swaps to the nearest neighbor.*

Before stating the assumption of convergence of the rescaled population at time 0, we clarify the space of measures and its topology. We will deal with random measures taking values in $\mathcal{M}_F(E)$, the space of finite positive measures on E, where mainly $E = \mathbb{R}_+$ or $E = \mathcal{X} \times \mathbb{R}_+$. The space $\mathcal{M}_F(E)$ can be embedded with the topology of the vague or weak convergence. Recall that the weak convergence in $\mathcal{M}_F(E)$ is defined as: $\nu_n \overset{weak}{\to} \nu$ if $\langle \nu_n, f\rangle \to \langle \nu, f\rangle$ for each continuous and bounded f. The vague convergence is defined as $\nu_n \overset{vague}{\to} \nu$ if $\langle \nu_n, f\rangle \to \langle \nu, f\rangle$ for each continuous f with compact support. In the case where E is a compact set, for example if a fixed limiting age is imposed, these topologies are the same, which contributes to simplify some assumptions (see Remark 1). In our case however, since $E = \mathcal{X} \times \mathbb{R}_+$ is not compact, these two topologies are strictly included.

In this paper, we denote $\mathbb{D}([0,T], (\mathcal{M}_F(E), w))$ (resp. $\mathcal{C}([0,T], (\mathcal{M}_F(E), w))$) the space of càdlàg (resp. continuous) processes on $[0,T]$ taking values in $\mathcal{M}_F(E)$ embedded with the topology of weak convergence.

The last assumption relates to the convergence of the sequence of initial populations towards a deterministic measure, assessing the repartition of ages and characteristics at initial time 0. This is stated below.

Assumption 4. *(Convergence of the initial population) There exists a deterministic measure $\tilde{\xi}_0 \in \mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+)$ such that the sequence of random measures $\tilde{Z}^n_0(\mathrm{d}x, \mathrm{d}a)$ converges in distribution and for the weak topology on $\mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+)$ to $\tilde{\xi}_0(\mathrm{d}x, \mathrm{d}a)$.*

Note that in the case of Assumption 4, we have in particular $\langle \tilde{Z}^n_0, \mathbf{1}\rangle = \frac{1}{n}\langle Z^n_0, \mathbf{1}\rangle$ converges to $\langle \xi_0, \mathbf{1}\rangle$ (in distribution thus in probability since the limit is deterministic), therefore the initial sample size $\langle Z^n_0, \mathbf{1}\rangle$ is assumed to be of order n.

2.3 Results

Here are detailed the main results of this paper. Proofs are given in Section 3. We before state the result on existence and strong uniqueness for the stochastic Equation (1). The reader is referred to Propositions 2.2.5 and 2.2.6 in Tran (2006) for the proof.

Proposition 1. *Under Assumptions 1 and 2, for each $n \in \mathbb{N}^*$ and $T > 0$, there exists a unique strong solution $\tilde{Z}^n_\cdot(\mathrm{d}x, \mathrm{d}a) \in \mathbb{D}([0,T], (\mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+), w))$ to Equation (1).*

Our main result is stated below.

Theorem 1. (i) *Under Assumptions 1, 2, 3 and 4, the sequence of age pyramid processes $(\bar{Z}^n_\cdot(\mathrm{d}a))_n$ (see Definition 1) converges in distribution in $\mathbb{D}([0,T], (\mathcal{M}_F(\mathbb{R}_+), w))$*

towards the unique (deterministic) measure-valued process $\xi \in \mathbb{C}([0,T], (\mathcal{M}_F(\mathbb{R}_+), w))$ *solution to: for each differentiable* f *on* $\mathbb{R}_+$*, with continuous derivatives,*

$$\langle \xi_t, f \rangle = \langle \xi_0, f \rangle + \int_0^t \langle \xi_s, \partial_a f + \hat{b} - \hat{d} \rangle \mathrm{d}s, \tag{3}$$

with initial condition $\xi_0(\mathrm{d}a) = \tilde{\xi}_0(\mathcal{X}, \mathrm{d}a)$*, and where*

$$\hat{b}(a) = \int_{\mathcal{X}} b(x,a) g(x,a) l(\mathrm{d}x), \ \hat{d}(a) = \int_{\mathcal{X}} d(x,a) g(x,a) l(\mathrm{d}x).$$

(ii) *For each* $t > 0$*, the sequence of random measures* $(\tilde{Z}^n_t(\mathrm{d}x, \mathrm{d}a))_n$ *converges in distribution in* $(\mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+), w)$ *to* $\tilde{\xi}_t(\mathrm{d}x, \mathrm{d}a) := g(x,a)\xi_t(\mathrm{d}a) l(\mathrm{d}x)$*.*

Equation (3) states that the macroscopic behavior of the age pyramid evolves as in a population with ages only, with birth and death rates given by $\hat{b}(a)$ and $\hat{d}(a)$. The weak formulation of Equation (3) is indeed the classical transport-renewal equation, known as McKendrick-Von Foerster,

$$(\partial_a + \partial_t)\xi_t(a) = -\hat{d}(a)\xi_t(a),$$
$$\xi_t(0) = \int_0^\infty \xi_t(a)\hat{b}(a)\mathrm{d}a.$$

In this model derived in point (i), the transport component states that each generation is aging and its size is decreased by the number of deaths over time, whereas the renewal component (initial condition in age) computes the number of newborns at time t based on the whole population alive at time t. Moreover, point (ii) claims first that the remaining time dependency only concerns the evolving age pyramid, and second that at each time t and age a, the characteristics distribution is given by $g(x,a)l(\mathrm{d}x)$. Note also that due to different time scales between birth and swap events, the mutation kernel at birth does not appear in the macroscopic dynamics. Let us emphasize that in the macroscopic model, the set of parameters is reduced: the birth, death and swap rates are "replaced" by birth and death rates which do not depend on characteristics anymore, driving the evolution of the age pyramid over time.

Let us describe the analogy with the work of Méléard and Tran (2012) in the following remark, which is used to establish our result.

Remark 4. *In Méléard and Tran (2012), the scaling limit of birth-death measure-valued processes is considered under some allometric component, that is additional fast births and deaths with same rate. The asymetry between birth (add an individual with age 0) and death events (remove an individual with positive age) leads to a major technical issue, and the problem can be tackled by assuming some fast aging* *phenomenon and extending averaging techniques by Kurtz (1992). In the limit, with fast birth and death events, the age pyramid reaches an equilibrium at each time. So, in the limit, the fast age component is stable and a Feller diffusion describes the evolution of the population structure in terms of characteristics which evolve at the slow time scale. The tools of Méléard and Tran (2012) are used in the proof of Theorem 1. In our framework including additional swap patterns, the situation is reversed: the aging component is at the slow time scale of the system, whereas characteristics evolve at the fast time scale. Analogously, in the limit, the characteristics structure is stable and a specific equation describes the evolution of the age pyramid. In our case also, the particularity of the swap phenomenon, which can be seen as special simultaneous birth and death events, leads to a limiting equation which is deterministic.*

A question which arises with point (ii) that sates a convergence for each fixed time, refers to the possibility to get the convergence in distribution of the whole sequence of processes. In fact, it is suggested in Méléard and Tran (2012) that the sequence of processes can not be tight. The following result is a statement about the non-tightness of the sequence of the measure-valued processes. This shows that for reasonable parameters, it is not possible to improve the result of Theorem 1. We state this results with characteristics only, and we make the following assumption:

Assumption 5. *Let us work with age-independent demographic rates and kernels, namely* $b(x,a) \equiv b(x)$, $d(x,a) \equiv d(x)$, $e(x,a) \equiv e(x)$, *and* $k_e(x,a,x') \equiv k_e(x,x')$. *We consider* Δ *to be the euclidian distance on* $\mathbb{R}^d$ *and for a given set* $A \subset \mathcal{X}$, *we denote* $\mathring{A}$ *its interior for the induced topology on* $\mathcal{X}$. *Let us assume that there exists two measurable non-empty and disjoint subsets* A *and* B *in* $\mathcal{X}$ *such that*
(i) $\Delta(A,B) = \inf_{x\in A, y\in B} \Delta(x,y) > 0$,
(ii) $\langle \xi_0, e\,\mathbf{1}_{\mathring{A}} \rangle > 0$, *where* $\mathbf{1}_{\mathring{A}}$ *is the indicator of the interior of* A,
(iii) $k_e(A,B) = \int_{x\in A}\int_{y\in B} k_e(x,y) m(\mathrm{d}x) m(\mathrm{d}y) > 0$.

These assumptions are quite natural: to sum up, they state that there exists some sub-population in the initial macroscopic population whose swap parameters allow to "escape" from the current characteristics. The non-tightness result is derived below.

Proposition 2. *Under Assumptions 4 and 5, the measure-valued process* $\tilde{Z}^n_{\cdot}(\mathrm{d}x, \mathrm{d}a)$ *is not tight in* $\mathbb{D}([0,T], \mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+))$.

3 Proofs

3.1 Proof of Theorem 1

Semi-martingale decomposition. The semi-martingale decomposition and the control of quadratic variations are key tools for limit theorems. In the following, decompositions are performed by compensation of the Poisson Point Measure, which give further insights on the behavior of the sequence of processes. In the following, we denote $C_b^{0,1}(\mathcal{X}\times\mathbb{R}_+)$ the set of bounded functions, continuous on $\mathcal{X}$ and differentiable on $\mathbb{R}_+$ with continuous and bounded partial derivative.

Lemma 1. *a) For each $f \in C_b^{0,1}(\mathcal{X}\times\mathbb{R}_+)$, the semi-martingale decomposition of the process $\langle \tilde{Z}_t^n, f\rangle$ is given by*

$$\langle \tilde{Z}_t^n, f\rangle = \langle \tilde{Z}_0^n, f\rangle + \int_0^t \langle \tilde{Z}_s^n,\, \partial_a f + H^{f,n}\rangle \mathrm{d}s + M_t^{f,n}, \tag{4}$$

where

$$\begin{aligned} H^{f,n}(x,a) = b(x,a)\int_{\mathcal{X}} f(x',0)k_b(x,a,x')l(\mathrm{d}x') - d(x,a)f(x,a) \\ + n\, e(x,a)\int_{\mathcal{X}} (f(x',a)-f(x,a))k_e(x,a,x')l(\mathrm{d}x') \end{aligned} \tag{5}$$

and $M^{f,n}$ is the local martingale (starting at zero) corresponding to the compensated Poisson point measure.

b) Let $\tau_k = \inf\{t:\ \langle \tilde{Z}_t^n, 1\rangle \geq k\}$. Then $M_{\cdot\wedge\tau_k}^{f,n}$ is a square-integrable martingale with quadratic variation

$$\begin{aligned} \langle M^{f,n}\rangle_{t\wedge\tau_k} = \frac{1}{n}\int_0^{t\wedge\tau_k} \mathrm{d}s \int_{\mathcal{X}} \tilde{Z}_s^n(\mathrm{d}x,\mathrm{d}a)\Big\{ b(x,a)\int_{\mathcal{X}} f(x',0)^2 k_b(x,a,x')l(\mathrm{d}x') \\ + d(x,a)f(x,a)^2 + n\, e(x,a)\int_{\mathcal{X}} (f(x',a)-f(x,a))^2 k_e(x,a,x')l(\mathrm{d}x')\Big\}. \end{aligned} \tag{6}$$

Proof of Lemma 1 The result is obtained by applying Equation (1) to a function $f \in C_b^{0,1}(\mathcal{X}\times\mathbb{R}_+)$. After integration by parts, the Poisson point measure Q can be compensated to get a local martingale, and the bracket can be computed. See Bensusan et al. (2010–2015) and Tran (2008) for more details. ⋄

As stated in Proposition 2, which is proved at the end of this section, the sequence of measure-valued processes $\tilde{Z}_\cdot^n(\mathrm{d}x,\mathrm{d}a)$ is not tight. This is in fact suggested by the term of order n in Equation (5). The remaining results we hope are
(i) the convergence for fixed t of the sequence $\tilde{Z}_t^n(\mathrm{d}x,\mathrm{d}a)$,
(ii) the convergence of the sequence of age pyramid processes $\bar{Z}_\cdot^n(\mathrm{d}a)$.

We now forget characteristics and focus on the age pyramid. By Lemma 1, the sequence of age pyramid processes verify for each bounded and differentiable $h(a)$,

$$\langle \bar{Z}^n_t, h\rangle = \langle \bar{Z}^n_0, h\rangle + \int_0^t \langle \bar{Z}^n_s,\, \partial_a h + h(0)b - h\, d\rangle \mathrm{d}s + M^{h,n}_t, \tag{7}$$

where $M^{h,n}_{.\wedge\tau_k}$ is a square-integrable martingale with quadratic variation

$$\langle M^{h,n}\rangle_{t\wedge\tau_k} = \frac{1}{n}\int_0^{t\wedge\tau_k} \langle \bar{Z}^n_s, h(0)^2 b + h^2\, d\rangle \mathrm{d}s. \tag{8}$$

Heuristically, the behavior of the quadratic variation in (8) is of order $\frac{1}{n}$, suggesting that the noise will vanish in the limit. Let us also remark the link, although not so obvious, between Equation (7) setting the noise to be zero, and the limit Equation (3) we want to derive.

Proof of Theorem 1. The proof is divided in seven steps. We follow the reasoning of Méléard and Tran (2012).

(i) Let us show that under Assumptions 1 and 2,

$$\sup_n \mathbb{E}\left[\sup_{t\le T}\langle \tilde{Z}^n_t, 1\rangle\right] < +\infty.$$

Recall that $\tau_k = \inf\{t:\ \langle \tilde{Z}^n_t, 1\rangle \ge k\}$. Equation (1) implies that

$$\langle \tilde{Z}^n_{t\wedge\tau_k}, 1\rangle \le \langle \tilde{Z}^n_0, \mathbf{1}\rangle + \frac{1}{n}\int_0^t\int_{N^*}\int_{\mathcal{X}}\int_{\mathbb{R}_+} \mathbf{1}_{i\le \sup_{0\le u\le s\wedge\tau_k}\langle Z^n_{s-},\mathbf{1}\rangle}\mathbf{1}_{0\le\theta<m_1(Z^n_{s-},i,x')} Q(\mathrm{d}s,\mathrm{d}i,\mathrm{d}x',\mathrm{d}\theta).$$

Take the supremum, then expectation and isometry formula, and finally use Assumption 1 to get

$$\mathbb{E}\left[\sup_{0\le u\le t}\langle \tilde{Z}^n_{t\wedge\tau_k}, 1\rangle\right] \le \mathbb{E}\left[\langle \tilde{Z}^n_0, \mathbf{1}\rangle\right] + \bar{b}\mathbb{E}\left[\int_0^t \sup_{0\le u\le s\wedge\tau_k}\langle \tilde{Z}^n_{s-}, \mathbf{1}\rangle \mathrm{d}s\right].$$

Fubini's theorem and Grönwall's lemma thus leads to

$$\mathbb{E}\left[\sup_{0\le u\le t}\langle \tilde{Z}^n_{t\wedge\tau_k}, 1\rangle\right] \le \mathbb{E}\left[\langle \tilde{Z}^n_0, \mathbf{1}\rangle\right]\exp(\bar{b}t).$$

Under Assumption 2, the right hand side is dominated by a constant which does not depend on n, that is for each n, $\mathbb{E}\left[\sup_{0\le u\le t}\langle \tilde{Z}^n_{t\wedge\tau_k}, 1\rangle\right] \le C\exp(\bar{b}t)$. This shows that $\tau_k \to +\infty$ a.s.. Then Fatou's lemma leads to $\mathbb{E}\left[\sup_{0\le u\le t}\langle \tilde{Z}^n_t, 1\rangle\right] \le C\exp(\bar{b}t)$. As the r.h.s. does not depend on n, we conclude that $\sup_n \mathbb{E}\left[\sup_{t\le T}\langle \tilde{Z}^n_t, 1\rangle\right] < +\infty$.

A direct corollary is that the local martingale $(M^{f,n}_t)$ with quadratic variation given by (6) is a square integrable martingale. Indeed,

$$\mathbb{E}\left[\langle M^{f,n}\rangle_{t\wedge\tau_k}\right] \le \|f\|^2_\infty\left(\bar{b}+\bar{d}+4\bar{e}\right)t\sup_n \mathbb{E}\left[\sup_{0\le s\le t}\langle \tilde{Z}^n_s, 1\rangle\right] < +\infty.$$

The last term being independent of k we get $\mathbb{E}\left[\langle M^{f,n}\rangle_t\right] < +\infty$ by Fatou's lemma, using that $\lim_{k\to+\infty}\tau_k = +\infty$.

A second corollary is that for fixed t the family of random variables $(M_t^{f,n})_n$ is uniformly integrable. Indeed, the previous inequality leads to

$$\sup_n \mathbb{E}\left[(M_t^{f,n})^2\right] = \sup_n \mathbb{E}\left[\langle M^{f,n}\rangle_t\right] \leq \|f\|_\infty^2 \left(\bar{b}+\bar{d}+4\bar{e}\right) t \sup_n \mathbb{E}\left[\sup_{0\leq s\leq t}\langle \tilde{Z}_s^n, 1\rangle\right] < +\infty.$$

(ii) We want to show that the sequence of age pyramid processes $\bar{Z}^n(\mathrm{d}a) = \int_{\mathcal{X}} \tilde{Z}^n(\mathrm{d}x, \mathrm{d}a)$ is tight in $\mathbb{D}\left([0,T], \mathcal{M}_F(\mathbb{R}_+)\right)$ where $\mathcal{M}_F(\mathbb{R}_+)$ is embedded with the topology of vague convergence. The extension to weak convergence is carried out in step (iii) .

According to Roelly-Coppoletta (1986) (Theorem 2.1), it is sufficient to prove that for each $f \in \Theta \cup \{\mathbf{1}\}$, $\langle \tilde{Z}^n, f\rangle$ is tight in $\mathbb{D}([0,T],\mathbb{R})$ where Θ is a dense subset of the space $C_0(\mathbb{R}_+,\mathbb{R})$, the space of continuous maps vanishing at infinity, for the topology of uniform convergence. By Tran (2006), Appendix A.2, the set $\Theta = C_0(\mathbb{R}_+,\mathbb{R}) \cap C_b^1(\mathbb{R}_+,\mathbb{R})$ is dense in $C_0(\mathbb{R}_+,\mathbb{R})$.

Let us show that for $f \in C_0(\mathbb{R}_+,\mathbb{R}) \cap C_b^1(\mathbb{R}_+,\mathbb{R})$, the sequence of $\mathbb{R}$-valued processes $\left(\langle \bar{Z}^n, f\rangle\right)_{n\in\mathbb{N}^*}$ is tight in $\mathbb{D}([0,T],\mathbb{R})$. To do this, we use Aldous-Rebolledo criterion (see Aldous (1978) and Joffe and Métivier (1986)):
since $\langle \bar{Z}_\cdot^n, f\rangle = V_\cdot^{n,f} + M_\cdot^{n,f}$ is a semi-martingale (with decomposition given in (7)), it sufficient to prove that

1. $\forall t \in [0,T]$, $\left(\langle M^{n,f}\rangle_t\right)_{n\in\mathbb{N}^*}$ and $\left(V_t^{n,f}\right)_{n\in\mathbb{N}^*}$ are uniformly tight in $\mathbb{R}$,
2. $\forall \epsilon > 0, \forall \eta > 0, \exists \delta > 0, \exists n_0 \in \mathbb{N}^*$ such that for any sequence of stopping times $(S_k)_{k\in\mathbb{N}^*}$ and $(T_k)_{k\in\mathbb{N}^*}$ verifying a.s. $\forall k \in \mathbb{N}^*, S_k \leq T_k \leq T$,

$$\begin{aligned} &\sup_{n\geq n_0} \mathbb{P}\left(\left|\langle M^{n,f}\rangle_{T_k} - \langle M^{n,f}\rangle_{S_k}\right| \geq \eta, T_k < S_k + \delta\right) \leq \epsilon, \\ &\sup_{n\geq n_0} \mathbb{P}\left(\left|V_{T_k}^{n,f} - V_{S_k}^{n,f}\right| \geq \eta, T_k < S_k + \delta\right) \leq \epsilon. \end{aligned} \tag{9}$$

To prove the first point, it is sufficient to show that

$$\sup_{n\in\mathbb{N}^*} \mathbb{E}\left[\sup_{0\leq t\leq T} \langle M^{n,f}\rangle_t\right] < +\infty \text{ and } \sup_{n\in\mathbb{N}^*} \mathbb{E}\left[\sup_{0\leq t\leq T}\left| V_t^{n,f}\right|\right] < +\infty.$$

According to (8) and step (i),

$$\sup_{n\in\mathbb{N}^*} \mathbb{E}\left[\sup_{0\leq t\leq T} \langle M^{n,f}\rangle_t\right] \leq \sup_{n\in\mathbb{N}^*} \mathbb{E}\left[\sup_{0\leq t\leq T} \langle \tilde{Z}_t^n, 1\rangle\right] T(\bar{b}+\bar{d}) \|f\|_\infty^2 < +\infty,$$

and from (7),

$$\begin{aligned} \sup_{n\in\mathbb{N}^*} \mathbb{E}\left[\sup_{0\leq t\leq T}\left| V_t^{n,f}\right|\right] &\leq \|f\|_\infty \sup_{n\in\mathbb{N}^*} \mathbb{E}\left[\langle \tilde{Z}_0^n, 1\rangle\right] \\ &\quad + \sup_{n\in\mathbb{N}^*} \mathbb{E}\left[\sup_{0\leq t\leq T}\langle \tilde{Z}^n, 1\rangle_t\right] T\left((\bar{b}+\bar{d})\|f\|_\infty + \|f'\|_\infty\right) < +\infty. \end{aligned}$$

To prove the second point (Equation (9)), let $\epsilon > 0, \eta > 0$ and $\forall k \in \mathbb{N}^*, S_k(\omega) \leq T_k(\omega) \leq T$ stopping times such that a.s. $S_k \leq T_k \leq S_k + \delta$. For all $n_0 \in \mathbb{N}^*$,

$$\begin{aligned}
&\sup_{n \geq n_0} \mathbb{P}\left(\left|\langle M^{n,f}\rangle_{T_k} - \langle M^{n,f}\rangle_{S_k}\right| \geq \eta\right) \\
&\leq \frac{1}{\eta} \sup_{n \geq n_0} \mathbb{E}\left[\left|\langle M^{n,f}\rangle_{T_k} - \langle M^{n,f}\rangle_{S_k}\right|\right] \\
&\leq \frac{\delta}{\eta} \|f\|_\infty^2 (\bar{b} + \bar{d}) \sup_{n \in \mathbb{N}^*} \mathbb{E}\left[\sup_{0 \leq t \leq T} \langle \tilde{Z}_t^n, 1\rangle\right] < +\infty.
\end{aligned}$$

Thus one can choose $\delta > 0$ such that

$$\sup_{n \geq n_0} \mathbb{P}\left(\left|\langle M^{n,f}\rangle_{T_k} - \langle M^{n,f}\rangle_{S_k}\right| \geq \eta\right) \leq \epsilon.$$

This shows that the sequence of processes $\langle \tilde{Z}^n, f\rangle$ is tight in $\mathbb{D}([0,T], \mathbb{R})$. Note that by the same reasoning, one can also show the tightness of $\langle \tilde{Z}^n, \mathbf{1}\rangle$. This concludes the proof of step (ii).

(iii) Let us now prove the tightness of $(\bar{Z}^n(\mathrm{d}a))_n$ in $\mathbb{D}([0,T], \mathcal{M}_F(\mathbb{R}_+))$ where $\mathcal{M}_F \equiv \mathcal{M}_F(\mathbb{R}_+)$ is embedded with the topology of weak convergence. To do this, we use the following criterion (see Méléard and Roelly (1993)). In the following, we denote $\Rightarrow$ the convergence in distribution of a sequence of processes.

Theorem 2. *Let (Z^n) be a sequence of processes in $\mathbb{D}([0,T], (\mathcal{M}_F, w))$ and ξ a process in the space $\mathbb{C}([0,T], (\mathcal{M}_F, w))$. Then the following statements are equivalent:*

- $Z^n \Rightarrow \xi$ *in* $\mathbb{D}([0,T], (\mathcal{M}_F, w))$,
- $Z^n \Rightarrow \xi$ *in* $\mathbb{D}([0,T], (\mathcal{M}_F, v))$ *and* $\langle Z^n, \mathbf{1}\rangle \Rightarrow \langle \xi, \mathbf{1}\rangle$ *in* $\mathbb{D}([0,T], \mathbb{R})$.

Recall that the tightness of $(\langle \bar{Z}^n, \mathbf{1}\rangle)$ in $\mathbb{D}([0,T], \mathbb{R})$ has been derived in the previous step. Now, since $(\bar{Z}^n(\mathrm{d}a))$ is tight in $\mathbb{D}([0,T], (\mathcal{M}_F, v))$ and $(\langle \bar{Z}^n, \mathbf{1}\rangle)$ is tight in $\mathbb{D}([0,T], \mathbb{R})$, then one can choose a subsequence $\phi(n)$ such that

$$\bar{Z}^{\phi(n)} \Rightarrow \xi^\phi \text{ in } \mathbb{D}([0,T], (\mathcal{M}_F, v)) \text{ and } \langle \bar{Z}^{\phi(n)}, \mathbf{1}\rangle \Rightarrow Y \text{ in } \mathbb{D}([0,T], \mathbb{R}),$$

where $(\xi_t^\phi)_{t \in [0,T]}$ is some process in $\mathbb{D}([0,T], (\mathcal{M}_F, v))$ and $(Y_t)_{t \in [0,T]}$ is some càdlàg real valued process. The aim now is to show that
a) $Y = \langle \xi^\phi, \mathbf{1}\rangle$, and
b) $\xi^\phi \in \mathbb{C}([0,T], (\mathcal{M}_F, w))$,
which will prove the convergence of $(\bar{Z}^{\phi(n)}(\mathrm{d}a))$ in $\mathbb{D}([0,T], (\mathcal{M}_F, w))$, with the help of the criterion in Theorem 2. This will show, by definition, that the sequence $(\bar{Z}^n(\mathrm{d}a))$ is tight in $\mathbb{D}([0,T], (\mathcal{M}_F, w))$.

a) Let us first prove that $\langle \bar{Z}^{\phi(n)}, \mathbf{1}\rangle \Rightarrow \langle \xi^{\phi}, \mathbf{1}\rangle$ in $\mathbb{D}\left([0,T],\mathbb{R}\right)$.

Let F Lipschitiz continuous and bounded function from $\mathbb{D}([0,T],\mathbb{R})$ to $\mathbb{R}$. Let us show that

$$\limsup_{n\to+\infty}\left|\mathbb{E}\left[F(\langle \tilde{Z}^{\phi(n)}, \mathbf{1}\rangle) - F(\langle \xi^{\phi}, \mathbf{1}\rangle)\right]\right| = 0.$$

To do this, we use the ideas of Jourdain et al. (2012) also used in Méléard and Tran (2012), in particular the following Lemma (see Lemma 4.3 in Jourdain et al. (2012)). The proof is postponed at the end of this section.

Lemma 2. *Introduce the functions* $f_k(a) := \Psi(0\vee(|a|-(k-1))\wedge 1)$*, where* $\Psi(y) = 6y^5 - 15y^4 + 10y^3$*, which are continuous approximations of the indicator function* $\mathbf{1}_{a\geq k}$*. Then under Assumptions of Theorem 1,*

$$\lim_{k\to+\infty}\limsup_{n\to+\infty}\mathbb{E}\left[\sup_{t\in[0,T]}\langle \tilde{Z}^{\phi(n)}, f_k\rangle\right] = 0,$$

and

$$\lim_{k\to+\infty}\mathbb{E}\left[\sup_{t\in[0,T]}\langle \xi^{\phi}_t, f_k\rangle\right] = 0.$$

Now, let us introduce the terms $F(\langle \tilde{Z}^{\phi(n)}, \mathbf{1}-f_k\rangle)$ and $F(\langle \xi^{\phi}, \mathbf{1}-f_k\rangle)$ to dominate by

$$\begin{aligned}
&\limsup_{n\to+\infty}\left|\mathbb{E}\left[F(\langle \tilde{Z}^{\phi(n)}, \mathbf{1}\rangle) - F(\langle \xi^{\phi}, \mathbf{1}\rangle)\right]\right| \\
&\leq \limsup_{k\to+\infty}\limsup_{n\to+\infty}\left|\mathbb{E}\left[F(\langle \tilde{Z}^{\phi(n)}, \mathbf{1}\rangle) - F(\langle \tilde{Z}^{\phi(n)}, \mathbf{1}-f_k\rangle)\right]\right| \\
&+ \limsup_{k\to+\infty}\limsup_{n\to+\infty}\left|\mathbb{E}\left[F(\langle \tilde{Z}^{\phi(n)}, \mathbf{1}-f_k\rangle) - F(\langle \xi^{\phi}, \mathbf{1}-f_k\rangle)\right]\right| \\
&+ \limsup_{k\to+\infty}\left|\mathbb{E}\left[F(\langle \xi^{\phi}, \mathbf{1}-f_k\rangle) - F(\langle \xi^{\phi}, \mathbf{1}\rangle)\right]\right|.
\end{aligned}$$

For the second term, since $\bar{Z}^{\phi(n)} \Rightarrow \xi^{\phi}$ in $\mathbb{D}\left([0,T],(\mathcal{M}_F,v)\right)$ and $(1-f_k)\in C_K(\mathbb{R}_+)$, the continuous mapping theorem (see e.g. Billingsley (2009)) implies that the second term is zero. Also, by the Lipschitz property,

$$|F(\langle \nu, \mathbf{1}-f_k\rangle) - F(\langle \nu, \mathbf{1}\rangle)| \leq [F]_{lip}\sup_{t\ \in[0,T]}\langle \nu_t, f_k\rangle,$$

so according to Lemma 2, the first term and third terms are equal to zero.

b) Let us now prove that $\xi^{\phi}\in\mathbb{C}\left([0,T],(\mathcal{M}_F,v)\right)$ (vague topology).
Since $\bar{Z}^{\phi(n)} \Rightarrow \xi$ in $\mathbb{D}\left([0,T],(\mathcal{M}_F,v)\right)$ and for each $f\in C_K(\mathbb{R}_+)$, the map $\nu\in D([0,T],(\mathcal{M}_F(\mathbb{R}_+),v))\mapsto \sup_{t\in[0,T]}|\langle \nu_t, f\rangle - \langle \nu_{t-}, f\rangle|$ is continuous, the continuous mapping theorem implies that

$$\sup_{t\in[0,T]}\left|\langle \bar{Z}^{\phi(n)}_t, f\rangle - \langle \bar{Z}^{\phi(n)}_{t-}, f\rangle\right| \Rightarrow \sup_{t\in[0,T]}\left|\langle \xi^{\phi}_t, f\rangle - \langle \xi^{\phi}_{t-}, f\rangle\right|.$$

Moreover, by construction if there is a jump the quantity $\langle \bar{Z}_t^{\phi(n)}, f\rangle$ increases at most by the amount $\frac{\|f\|_\infty}{\Phi(n)}$, which turns out that $\sup_{t\in[0,T]}\left|\langle \bar{Z}_t^{\phi(n)}, f\rangle - \langle \bar{Z}_{t-}^{\phi(n)}, f\rangle\right| \leq \frac{\|f\|_\infty}{\Phi(n)}$, and in particular $\sup_{t\in[0,T]}\left|\langle \bar{Z}_t^{\phi(n)}, f\rangle - \langle \bar{Z}_{t-}^{\phi(n)}, f\rangle\right| \Rightarrow 0$. This shows that for each $f \in C_K(\mathbb{R}_+)$, a.s. $\sup_{t\in[0,T]}\left|\langle \xi_t^{\phi}, f\rangle - \langle \xi_{t-}^{\phi}, f\rangle\right| = 0$ thus $\xi^\phi \in \mathbb{C}\left([0,T], (\mathcal{M}_F, v)\right)$.

Now, we prove that $\xi^\phi \in \mathbb{C}\left([0,T], (\mathcal{M}_F, w)\right)$ (weak topology). Let $h \in C_b(\mathbb{R}_+)$ and the (f_k) as in Lemma 2. Then

$$\begin{aligned}\mathbb{E}\left[\sup_{t\in[0,T]}\left|\langle \xi_t^{\phi}, h\rangle - \langle \xi_{t-}^{\phi}, h\rangle\right|\right] \leq \|h\|_\infty \mathbb{E}\left[\sup_{t\in[0,T]}\left|\langle \xi_t^{\phi}, f_k\rangle\right|\right] \\ + \mathbb{E}\left[\sup_{t\in[0,T]}\left|\langle \xi_t^{\phi}, h(\mathbf{1}-f_k)\rangle - \langle \xi_{t-}^{\phi}, h(\mathbf{1}-f_k)\rangle\right|\right] \\ + \|h\|_\infty \mathbb{E}\left[\sup_{t\in[0,T]}\left|\langle \xi_{t-}^{\phi}, f_k\rangle\right|\right].\end{aligned}$$

Since $\xi^\phi \in \mathbb{C}\left([0,T], (\mathcal{M}_F, v)\right)$ (vague topology) and $h(\mathbf{1}-f_k) \in C_K(\mathbb{R}_+)$, the second term is zero. In addition, by the use of Lemma 2, the first and third terms are zero by letting $k \to +\infty$.

(iv) Let us prove that for each $t > 0$, the sequence of measures $(\tilde{Z}_t^n(\mathrm{d}x, \mathrm{d}a))_n$ is uniformly tight in $\mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+)$.

To show the tightness of a sequence of measures, we use the following result (see Kurtz (1992), Lemma 1.1):

Lemma 3. *Let (μ_n) be a sequence of random variables with values in $(\mathcal{M}_F(S), w)$, with (S, d) a complete separable metric space. Then (μ_n) is relatively compact in $(\mathcal{M}_F(S), w)$ if and only if the sequence $(\mu_n(S))$ is relatively compact in $\mathbb{R}$ and for each $\epsilon > 0$ there exists a compact set $K \subset S$ such that $\sup_n \mathbb{P}\left(\mu_n(K^c) > \epsilon\right) < \epsilon$.*

First note that $\tilde{Z}_t^n(\mathcal{X} \times \mathbb{R}_+) = \langle \tilde{Z}_t^n, \mathbf{1}\rangle$. Let $\epsilon > 0$, then for k large enough,

$$\sup_{n\geq 1} \mathbb{P}(\langle \tilde{Z}_t^n, \mathbf{1}\rangle \notin [0,k]) \leq \frac{1}{k} \sup_{n\geq 1} \mathbb{E}[\langle \tilde{Z}_t^n, \mathbf{1}\rangle] \leq \epsilon,$$

so that the first condition is matched. For the second condition, let A>0 and consider the compact set $K := \mathcal{X} \times [0, A]$. Obviously, $\tilde{Z}_t^n(K^c) = 0$ since all characteristics lie in $\mathcal{X}$: this proves that the second condition is satisfied.

The tightness of the sequence of the age pyramid processes and the time marginal of the population process has been established. To prove the convergence in distribution, it is needed to identify the limiting values. In our framework, this issue is technical due to the fast swap pattern: it is not possible to keep track of the characteristics in the population over time. The issue has an analogy with that in

Méléard and Tran (2012) and Gupta et al. (2014) in which fast aging is considered (see Remark 4). The main technical tool they use for the proof is an extension of averaging techniques by Kurtz (1992). This is used in the following step: the idea is to identify the limiting values of the sequence of occupation measures $(\Gamma^n(\mathrm{d}x, \mathrm{d}a, \mathrm{d}t))_n := (\tilde{Z}_t^n(\mathrm{d}x, \mathrm{d}a)\mathrm{d}t)_n$.

(v) To show that the sequence of measures $(\Gamma^n(\mathrm{d}x, \mathrm{d}a, \mathrm{d}t))_n$ is tight in $(\mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+ \times [0,T]), w)$, we use the same result as for the previous step. As previously, the second condition is straightforward with $\tilde{K} := \mathcal{X} \times [0, A] \times [0,T]$. As for the first condition, we have

$$\sup_n \mathbb{P}(\Gamma^n(\mathcal{X}\times\mathbb{R}_+\times[0,T]) \notin [0,k]) = \sup_n \mathbb{P}\left(\int_0^T \langle \tilde{Z}_t^n, \mathbf{1}\rangle \mathrm{d}t > k\right) \leq \frac{T}{k} \sup_n \mathbb{E}\left[\sup_{0\leq t\leq T} \langle \tilde{Z}_t^n, \mathbf{1}\rangle\right],$$

which can be made smaller than ϵ for k large enough.

Now, from (iv) and (v), the sequence $(\Gamma^n, \bar{Z}_\cdot^n(\mathrm{d}a))_n$ is tight. Due to Prohorov theorem, from any subsequence, one can extract a further sub-subsequence converging in distribution. Denote $(\Gamma^{\Phi(n)}, \bar{Z}_\cdot^{\Phi(n)}(\mathrm{d}a))$ such sub-subsequence and $(\bar{\Gamma}^\Phi, \bar{X}_\cdot^\Phi(\mathrm{d}a))$ the corresponding limit (the first component being a measure and the second component a measure-valued process). The last two steps of the proof are dedicated to the characterization of this limit. If it is unique (i.e. does not depend on Φ), we get the convergence in distribution (see Billingsley (2009), Theorem 2.6). Step (vi) focuses on the marginal in characteristics of $\bar{\Gamma}^\Phi$, whereas Step (vii) concentrates on the age and time marginal.

(vi) By definition of $\Gamma^{\Phi(n)}$, for each continuous and bounded $f(a, s)$,

$$\int_0^t \int_{\mathbb{R}_+} \int_{\mathcal{X}} f(a,s)\Gamma^{\Phi(n)}(\mathrm{d}x, \mathrm{d}a, \mathrm{d}s) = \int_0^t \int_{\mathbb{R}_+} f(a,s)\bar{Z}_s^{\Phi(n)}(\mathrm{d}a)\mathrm{d}s.$$

Since $\Gamma \mapsto \int_0^t \int_{\mathbb{R}_+} \int_{\mathcal{X}} f(a,s)\Gamma(\mathrm{d}x, \mathrm{d}a, \mathrm{d}s)$ and $Z \mapsto \int_0^t \int_{\mathbb{R}_+} f(a,s)Z_s(\mathrm{d}a)\mathrm{d}s$ are both continuous, the continuous mapping theorem implies that $\bar{X}_s^\Phi(\mathrm{d}a)\mathrm{d}s$ is necessarily the marginal measure of $\bar{\Gamma}^\Phi$ on $\mathbb{R}_+ \times [0,T]$ up to a null-measure set.

By Kurtz (1992), Lemma 1.4, there exists a predictable probability-valued process $\gamma_{a,s}^\Phi(\mathrm{d}x)$ such that a.s., $\mathrm{d}t - ae$, $\bar{\Gamma}^\Phi(\mathrm{d}x, \mathrm{d}a, \mathrm{d}s) = \gamma_{a,s}^\Phi(\mathrm{d}x)\bar{X}_s^\Phi(\mathrm{d}a)\mathrm{d}s$.

Now, the aim is to characterize $\gamma_{a,s}^\Phi(\mathrm{d}x)$. From (4), $(\frac{1}{\Phi(n)}M_t^{f,\Phi(n)})_t$ is a martingale, converging in distribution to (continuous mapping theorem for $(\Gamma^{\Phi(n)})_n$)

$$\bar{M}_t^\Phi := \int_0^t \int_{\mathbb{R}_+} \int_{x\in\mathcal{X}} e(x,a)\left(\int_{\mathcal{X}} (f(x',a) - f(x,a))k_e(x,a,x')l(\mathrm{d}x')\right) \gamma_{a,s}^\Phi(\mathrm{d}x)\bar{X}_s^\Phi(\mathrm{d}a)\mathrm{d}s.$$

For a given $t > 0$ the family $\left(\frac{1}{n}M_t^{f,n}\right)_n$ is uniformly integrable since from (i),

$$\sup_n \mathbb{E}\left[(M_t^{f,n})^2\right] = \sup_n \mathbb{E}\left[\langle M^{f,n}\rangle_t\right] \leq \|f\|_\infty^2 \left(\bar{b} + \bar{d} + 4\bar{e}\right) t \sup_n \mathbb{E}\left[\sup_{0\leq s\leq t} \langle \tilde{Z}_s^n, 1\rangle\right] < +\infty.$$

This implies that $\bar{M}^\Phi$ is a martingale. But it is by construction a finite-variation process which is also continuous so it is the null process (up to indistinguishability). We get *a.s.*, $X_s^\Phi(\mathrm{d}a) - a.e.$ and $\mathrm{d}t - a.e.$,

$$\int_{x\in\mathcal{X}} f(x,a)e(x,a)\gamma_{a,s}^\Phi(\mathrm{d}x) = \int_{x'\in\mathcal{X}} f(x',a)\left(\int_{x\in X} e(x,a)k_e(x,a,x')\gamma_{a,s}^\Phi(\mathrm{d}x)\right) l(\mathrm{d}x').$$

This shows that ($a.s.$, $X_s^\Phi(\mathrm{d}a) - a.e.$ and $\mathrm{d}t - a.e.$) $\gamma_{a,s}^\Phi$ is absolutely continuous w.r.t. l, $\gamma_{a,s}^\Phi(\mathrm{d}x) = \gamma_{a,s}^\Phi(x)l(\mathrm{d}x)$, with

$$\gamma_{a,s}^\Phi(x)e(x,a) = \int_{y\in\mathcal{X}} e(y,a)k_e(y,a,x)\gamma_{a,s}^\Phi(y)l(\mathrm{d}y),$$

and in addition γ_s^Φ is a probability measure. Then under Assumption 3, we get $\gamma_{a,s}^\Phi(x) = g(x,a)$. We just characterized that all limiting values of $\Gamma^{\Phi(n)}$ are of the form $\bar{\Gamma}^\Phi(\mathrm{d}x, \mathrm{d}a, \mathrm{d}s) = g(x,a)l(\mathrm{d}x)\bar{X}_s^\Phi(\mathrm{d}a)\mathrm{d}s$. In the last step, we identify $\bar{X}^\Phi$ as the solution to a deterministic equation.

(vii) We want to show that a.s. for each t and $f \in C_b^1(\mathbb{R}_+)$,

$$\langle \bar{X}_t^\Phi, f\rangle = \langle \xi_0, f\rangle + \int_0^t \langle \bar{X}_s^\Phi,\, \partial_a f + \hat{b} - \hat{d}\rangle \mathrm{d}s,$$

where $\hat{b}(a) = \int_\mathcal{X} b(x,a)g(x,a)l(\mathrm{d}x)$ and $\hat{d}(a) = \int_\mathcal{X} d(x,a)g(x,a)l(\mathrm{d}x)$. Let $\bar{H}_t^\Phi = \langle \bar{X}_t^\Phi, f\rangle_t - \langle \xi_0, f\rangle - \int_0^t \langle \bar{X}_s^\Phi,\, \partial_a f + \hat{b} - \hat{d}\rangle \mathrm{d}s$. To prove that $\mathbb{E}\left[\left|\bar{H}_t^\Phi\right|\right] = 0$, we use the following three facts:

a) From Equation (7), one has $M_t^{f,\Phi(n)} \Rightarrow \bar{H}_t^\Phi$, using that $\Gamma^{\Phi(n)} \Rightarrow \bar{\Gamma}^\Phi$, Assumption 4 and the continuous mapping theorem.

b) One can also write

$$\begin{aligned}\mathbb{E}\left[\left|M_t^{f,\Phi(n)}\right|\right]^2 &\leq \mathbb{E}\left[\left(M_t^{f,\Phi(n)}\right)^2\right] = \mathbb{E}\left[\langle M^{f,\Phi(n)}\rangle_t\right] \\ &\leq \frac{1}{\Phi(n)} t\,\|f\|_\infty^2 \left(\bar{b} + \bar{d}\right) \sup_{n\geq 1} \mathbb{E}\left[\sup_{0\leq s\leq t} \langle \tilde{Z}_s^n, \mathbf{1}\rangle\right],\end{aligned}$$

which shows that $\lim_{n\to+\infty} \mathbb{E}\left[\left|M_t^{f,\Phi(n)}\right|\right] = 0$ since $\Phi(n) \to +\infty$.

c) From the second corollary of (i), for fixed t, the sequence $(M_t^{f,\Phi(n)})_n$ is uniformly integrable.

From these facts we get $\mathbb{E}\left[\left|\bar{H}_t^\Phi\right|\right] = \lim_{n\to+\infty} \mathbb{E}\left[\left|M_t^{f,\Phi(n)}\right|\right] = 0$.

Using Grönwall lemma, one can prove the uniqueness for Equation (3). This concludes the proof.

Proof of Lemma 2. The proof is similar to Jourdain et al. (2012), exposed in order to illustrate that here only the moments of order $1+\alpha$ of Assumption 2 are needed (order 1 suffices elsewhere in the proof of Theorem 1). Note that since $f_k \notin C_K(\mathbb{R}_+)$, it is not possible to use the fact that $(\bar{Z}^{\Phi(n)}(\mathrm{d}a))$ converges in distribution in $\mathbb{D}([0,T],(\mathcal{M}_F,v))$. We rather use Equation (7) with $h \equiv f_k$ for $k \geq 1$ leads to (since $f_k(0)=0$),

$$\langle \bar{Z}_t^{\Phi(n)}, f_k \rangle = \langle \bar{Z}_0^{\Phi(n)}, f_k \rangle + \int_0^t \langle \bar{Z}_s^{\Phi(n)}, \partial_a f_k - f_k\, d \rangle \mathrm{d}s + M_t^{f_k,\Phi(n)}. \tag{10}$$

Dominate by omitting the death term, take the supremum, then expectation and finally use Doob inequality to get

$$\mathbb{E}\left[\sup_{t\in[0,T]} \langle \bar{Z}_t^{\Phi(n)}, f_k \rangle\right] \leq \mathbb{E}\left[\langle \bar{Z}_0^{\Phi(n)}, f_k \rangle\right] + 2\left(\mathbb{E}\left[\langle M^{f_k,\Phi(n)}\rangle_T\right]\right)^{1/2} + \int_0^T \mathbb{E}\left[\langle \bar{Z}_s^{\Phi(n)}, \partial_a f_k \rangle\right] \mathrm{d}s.$$

The limits of the three terms are studied separately.

(i) Since $\tilde{Z}_0^{\phi(n)} \Rightarrow \xi_0$ in $(\mathcal{M}_F(\mathcal{X}\times\mathbb{R}_+), w)$ and $f_k \in C_b(\mathcal{X}\times\mathbb{R}_+)$, then $\langle \bar{Z}_0^{\Phi(n)}, f_k \rangle \Rightarrow \langle \xi_0, f_k \rangle$. We also have $f_k \leq \mathbf{1}$ and under Assumption 2, $\sup_n \mathbb{E}\left[\left(\langle \bar{Z}_0^n, \mathbf{1}\rangle\right)^{1+\alpha}\right] < +\infty$ so the sequence is uniformly integrable. We thus get $\lim_{n\to+\infty} \mathbb{E}\left[\langle \bar{Z}_0^{\Phi(n)}, f_k \rangle\right] = \langle \xi_0, f_k \rangle$. Finally, since $\xi_0 \in \mathcal{M}_F(\mathcal{X}\times\mathbb{R}_+)$, that is $\langle \xi_0, \mathbf{1}\rangle < +\infty$, we get $\lim_{k\to+\infty} \langle \xi_0, f_k\rangle = 0$, then

$$\lim_{k\to+\infty} \limsup_{n\to+\infty} \mathbb{E}\left[\langle \bar{Z}_0^{\Phi(n)}, f_k \rangle\right] = 0.$$

(ii) From (8), since $f_k(0)=0$,

$$\mathbb{E}\left[\langle M^{f_k,\Phi(n)}\rangle_T\right] = \frac{1}{\Phi(n)} \int_0^t \langle \tilde{Z}_s^{\Phi(n)}, f_k^2\, d \rangle \mathrm{d}s \leq \frac{\bar{d}T}{\Phi(n)} \sup_{n\geq 1} \mathbb{E}\left[\sup_{t\in[0,T]} \langle \tilde{Z}_s^n, \mathbf{1}\rangle\right],$$

which shows that $\lim_{k\to+\infty} \limsup_{n\to+\infty} \mathbb{E}\left[\langle M^{f_k,\Phi(n)}\rangle_T\right] = 0$.

(iii) Note that

$$\partial_a f_k = \mathbf{1}_{k-1\leq a\leq k} \Psi'(a-(k-1)) \leq \mathbf{1}_{k-1\leq a \leq k} \sup_{y\in[0,1]} \Psi'(y) \leq \mathbf{1}_{k-1\leq a} \sup_{y\in[0,1]} \Psi'(y).$$

Since $\mathbf{1}_{k-1\leq a} \leq f_{k-1}(a)$, this shows that $\partial_a f_k(a) \leq f_{k-1}(a) \sup_{y\in[0,1]} \Psi'(y)$. To show that $\lim_{k\to+\infty} \limsup_{n\to+\infty} \int_0^T \mathbb{E}\left[\langle \bar{Z}_s^{\Phi(n)}, \partial_a f_k \rangle\right] \mathrm{d}s = 0$, it is thus sufficient to prove that for $t\in[0,T]$, $\lim_{k\to+\infty} \limsup_{n\to+\infty} \mathbb{E}\left[\langle \bar{Z}_t^{\Phi(n)}, f_k \rangle\right] = 0$. From (10), for each $k \geq 1$, with $C := \sup_{y\in[0,1]} \Psi'(y)$,

$$\mathbb{E}\left[\langle \bar{Z}_t^{\Phi(n)}, f_k \rangle\right] \leq \mathbb{E}\left[\langle \bar{Z}_0^{\Phi(n)}, f_k \rangle\right] + C \int_0^t \mathbb{E}\left[\langle \bar{Z}_s^{\Phi(n)}, f_{k-1} \rangle\right] \mathrm{d}s, \tag{11}$$

and for $k=0$, $f_0 \equiv \mathbf{1}$ then

$$\mathbb{E}\left[\langle \bar{Z}_t^{\Phi(n)}, f_0 \rangle\right] \leq \mathbb{E}\left[\langle \bar{Z}_0^{\Phi(n)}, f_0 \rangle\right] + \bar{b} \int_0^t \mathbb{E}\left[\langle \bar{Z}_s^{\Phi(n)}, f_0 \rangle\right] \mathrm{d}s,$$

so Grönwall's lemma leads to

$$\mathbb{E}\left[\langle \bar{Z}_t^{\Phi(n)}, f_0\rangle\right] \leq \mathbb{E}\left[\langle \bar{Z}_0^{\Phi(n)}, f_0\rangle\right] \exp(\bar{b}t). \tag{12}$$

From (11) and (12) we get

$$\mathbb{E}\left[\langle \bar{Z}_t^{\Phi(n)}, f_k\rangle\right] \leq \sum_{j=0}^{k-1} \frac{C^j t^j}{j!} \mathbb{E}\left[\langle \bar{Z}_0^{\Phi(n)}, f_{k-j}\rangle\right] + \frac{C^k t^k}{k!} \exp(\bar{b}t)\mathbb{E}\left[\langle \bar{Z}_0^{\Phi(n)}, \mathbf{1}\rangle\right].$$

Under Assumption 2, we have $\frac{C^k t^k}{k!} \exp(\bar{b}t)\mathbb{E}\left[\langle \bar{Z}_0^{\Phi(n)}, \mathbf{1}\rangle\right] \leq \frac{C^k t^k}{k!} \exp(\bar{b}t) \sup_n \mathbb{E}\left[\langle \bar{Z}_0^{n}, \mathbf{1}\rangle\right]$, which tends to zero as $k \to +\infty$. As for the first term, this can be split into

$$\begin{aligned}\sum_{j=0}^{k-1} \frac{C^j t^j}{j!} \mathbb{E}\left[\langle \bar{Z}_0^{\Phi(n)}, f_{k-j}\rangle\right] &= \sum_{j=0}^{[k/2]} \frac{C^j t^j}{j!} \mathbb{E}\left[\langle \bar{Z}_0^{\Phi(n)}, f_{k-j}\rangle\right] + \sum_{j=[k/2]+1}^{k-1} \frac{C^j t^j}{j!} \mathbb{E}\left[\langle \bar{Z}_0^{\Phi(n)}, f_{k-j}\rangle\right] \\ &\leq \exp(Ct)\mathbb{E}\left[\langle \bar{Z}_0^{\Phi(n)}, f_{k-[k/2]}\rangle\right] + \mathbb{E}\left[\langle \bar{Z}_0^{\Phi(n)}, \mathbf{1}\rangle\right] \sum_{j=[k/2]+1}^{k-1} \frac{C^j t^j}{j!}.\end{aligned}$$

Since $[k/2] \leq k - [k/2]$ then $f_{k-[k/2]} \leq f_{[k/2]}$. As in (i), by the convergence in distribution and uniform integrability (Assumption 2), one gets $\lim_{k\to+\infty} \limsup_{n\to+\infty} \mathbb{E}\left[\langle \bar{Z}_0^{\Phi(n)}, f_{[k/2]}\rangle\right] = 0$. As for the second term, this converges to zero since the first component is bounded (Assumption 2) and the sum converges to zero. $\diamond$

3.2 Proof of Proposition 2

Let us recall below the assumptions and the corresponding result about the non-tightness of the sequence of population processes.

Assumption 5 *Let us work with age-independent demographic rates and kernels, namely* $b(x, a) \equiv b(x)$, $d(x, a) \equiv d(x)$, $e(x, a) \equiv e(x)$, *and* $k_e(x, a, x') \equiv k_e(x, x')$. *We consider* Δ *to be the euclidian distance on* $\mathbb{R}^d$ *and for a given set* $A \subset \mathcal{X}$, *we denote* $\mathring{A}$ *its interior for the induced topology on* $\mathcal{X}$. *Let us assume that there exists two measurable non-empty and disjoint subsets* A *and* B *in* $\mathcal{X}$ *such that*
(i) $\Delta(A, B) = \inf_{x\in A, y\in B} \Delta(x, y) > 0$,
(ii) $\langle \xi_0, e\,\mathbf{1}_{\mathring{A}}\rangle > 0$, *where* $\mathbf{1}_{\mathring{A}}$ *is the indicator of the interior of* A,
(iii) $k_e(A, B) = \int_{x\in A}\int_{y\in B} k_e(x, y)m(\mathrm{d}x)m(\mathrm{d}y) > 0$.

Proposition 2 *Under Assumptions 4 and 5, the measure-valued process* $\tilde{Z}^n_{\cdot}(\mathrm{d}x, \mathrm{d}a)$ *is not tight in* $\mathbb{D}([0, T], \mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+))$.

Proof of Proposition 2

Aim is to give here a proof which does not need the main result of Theorem 1. Let us first construct a function $f \in \mathcal{C}_b(\mathcal{X})$ such that

$$\inf_{x\in A, y\in B} |f(y) - f(x)| > 0.$$

To this aim, consider for example the map $f : x \mapsto \Delta(x, A)$ with support in $\mathcal{X}\backslash A$, that satisfies the previous equation according to point (i). It is moreover continuous on the compact $\mathcal{X}$ so it is bounded. To prove that the sequence of measure-valued processes $(\tilde{Z}^n_\cdot(\mathrm{d}x, \mathrm{d}a))_n$ is not tight in $\mathbb{D}([0,T], \mathcal{M}_F(\mathcal{X} \times \mathbb{R}_+))$, it is sufficient to prove that the process $X^n_\cdot := \langle \tilde{Z}^n_\cdot, f\rangle$ is not tight in $\mathbb{D}([0,T], \mathbb{R})$. Recall the following equivalence derived from compactness characterization in the space of càdlàg functions (see Billingsley (2009), Theorem 13.2):

The sequence $(X^n_\cdot)_n$ is tight $\mathbb{D}([0,T], \mathbb{R})$ if and only if

a) the sequence $\sup_{0\leq s\leq T} |X^n_t|$ is tight in $\mathbb{R}$ and

b) (equicontinuity) $\forall \eta > 0$, $\lim_{\delta\to 0} \limsup_{n\to+\infty} \mathbb{P}\left(w'\left(X^n, \delta\right) > \eta\right) = 0$, where $w'\left(X^n, \delta\right) = \inf_{(t_i)>\delta} \sup_i \sup_{s,t\in[t_{i-1},t_i[} |X^n_t - X^n_s|$ and $\inf_{(t_i)>\delta}$ denotes the infimum over all subdivisions of $[0,T]$ with step strictly greater that δ.

To prove the result, we show that the equicontinuity criterion is not satisfied, that is we show that:

$$\exists\eta > 0, \exists\epsilon > 0, \forall\delta > 0, \limsup_{n\to+\infty} \mathbb{P}\left(\inf_{(t_i)>\delta} \sup_i \sup_{s,t\in[t_{i-1},t_i[} |X^n_t - X^n_s| > \eta\right) > \epsilon.$$

First remark that for each $\delta > 0$ and $\eta > 0$,

$$\mathbb{P}\left(\inf_{(t_i)>\delta} \sup_i \sup_{s,t\in[t_{i-1},t_i[} |X^n_t - X^n_s| > \eta\right) \geq \mathbb{P}\left(\sup_{t\in[0,\delta]} \left|X^n_t - X^n_{t-}\right| > \eta\right).$$

Second, denote τ^n the first time of jump of $Z^n_\cdot$. Conditionally on Z^n_0, τ^n is exponentially distributed with parameter $n\langle Z^n_0, e\rangle + \langle Z^n_0, b\rangle + \langle Z^n_0, d\rangle$. Then

$$\mathbb{P}\left(\tau^n \leq \delta \mid Z^n_0\right) = 1 - e^{-\delta\left(n\langle Z^n_0, e\rangle + \langle Z^n_0, b\rangle + \langle Z^n_0, d\rangle\right)}.$$

Let τ^n_e be the first time of swap for $Z^n_\cdot$. By independence between swap, birth and death events, the probability that the first time of event is a swap is given by (with convention $\frac{0}{0} = 0$)

$$\mathbb{P}\left(\tau^n = \tau^n_e \mid Z^n_0\right) = \frac{n\langle Z^n_0, e\rangle}{n\langle Z^n_0, e\rangle + \langle Z^n_0, b\rangle + \langle Z^n_0, d\rangle},$$

and the clock lemma for independent exponentially distributed random variables assesses independence, conditionally on Z^n_0, between the events $\{\tau^n = \tau^n_e\}$ and $\{\tau^n \leq \delta\}$.

Let Y be the random variable "characteristics of the individual for which the event occurs at time τ^n", and A and B as in Assumption 5. The probability that the individual has characteristics in A conditionally on a swap event is

$$\mathbb{P}\left(Y \in A \mid Z_0^n, \tau^n = \tau_e^n\right) = \frac{\langle Z_0^n, e\mathbf{1}_A\rangle}{\langle Z_0^n, e\rangle}.$$

Denote Y' the new characteristics (at birth or swap). In case of a swap, the characteristics Y are replaced by the new characteristics Y', drawn with distribution $k_e(Y,.)$. Then

$$\mathbb{P}\left(Y' \in B \mid Y \in A, \tau^n = \tau_e^n, Z_0^n\right) = \frac{k_e(A,B)}{m(A)}.$$

Finally, remark that Y and Y' are independent of τ^n and let us take $\eta = \frac{1}{2}\inf_{x\in A, y\in B}|f(y)-f(x)|$ which is positive by construction of f. Then

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\in[0,\delta]}\left|X_t^n - X_{t-}^n\right| > \eta\right)\\
&\geq \mathbb{P}\left(|X_{\tau^n\wedge\delta}^n - X_0^n| > \eta\right)\\
&\geq \mathbb{P}\left(\tau^n \leq \delta, \tau^n = \tau_e^n, Y \in A, Y' \in B\right)\\
&= \mathbb{E}\left[\mathbb{P}\left(\tau^n \leq \delta \mid Z_0^n\right)\mathbb{P}\left(\tau^n = \tau_e^n, Y' \in B, Y \in A \mid Z_0^n\right)\right]\\
&= \mathbb{E}\left[\mathbb{P}\left(\tau^n \leq \delta \mid Z_0^n\right)\mathbb{P}\left(Y' \in B \mid \tau^n = \tau_e^n, Y \in A, Z_0^n\right)\mathbb{P}\left(Y \in A \mid \tau^n = \tau_e^n, Z_0^n\right)\mathbb{P}\left(\tau^n = \tau_e^n \mid Z_0^n\right)\right]\\
&= \frac{k_e(A,B)}{m(A)}\mathbb{E}\left[\left(1 - e^{-\delta\left(n\langle Z_0^n,e\rangle + \langle Z_0^n,b\rangle + \langle Z_0^n,d\rangle\right)}\right)\frac{\langle Z_0^n, e\mathbf{1}_A\rangle}{\langle Z_0^n, e\rangle}\frac{n\langle Z_0^n, e\rangle}{n\langle Z_0^n, e\rangle + \langle Z_0^n, b\rangle + \langle Z_0^n, d\rangle}\right]
\end{aligned}$$

where the last two equalities are due to recursive conditioning and the use of the previous computations respectively. Now, let us construct $h_A \in \mathcal{C}_b(\mathcal{X})$ such that $h_A \leq \mathbf{1}_A$ and $\langle \xi_0, eh_A\rangle > 0$. Let us define

$$h_A : x \mapsto \frac{\Delta(x, \mathcal{X}\backslash A)}{\sup_{y\in A}\Delta(y, \mathcal{X}\backslash A)},$$

which is continuous on $\mathcal{X}$ and positive on $\mathring{A}$; according to point (ii) in Assumption 5, we get $\langle \xi_0, eh_A\rangle > 0$. Since $\langle \xi_0, eh_A\rangle \leq \langle \xi_0, e\mathbf{1}_A\rangle$, we have

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\in[0,\delta]}\left|X_t^n - X_{t-}^n\right| > \eta\right)\\
&\geq \frac{k_e(A,B)}{m(A)}\mathbb{E}\left[\left(1 - e^{-\delta\left(n\langle Z_0^n,e\rangle + \langle Z_0^n,b\rangle + \langle Z_0^n,d\rangle\right)}\right)\frac{n\langle Z_0^n, eh_A\rangle}{n\langle Z_0^n, e\rangle + \langle Z_0^n, b\rangle + \langle Z_0^n, d\rangle}\right]\\
&\geq \frac{k_e(A,B)}{m(A)}\mathbb{E}\left[\frac{n\delta\langle Z_0^n, eh_A\rangle}{1 + \delta\left(n\langle Z_0^n, e\rangle + \langle Z_0^n, b\rangle + \langle Z_0^n, d\rangle\right)}\right]
\end{aligned}$$

where the last inequality comes from the fact that $1 - \exp(-x) \geq \frac{x}{1+x}$. Renormalization by δn^2 leads to the final inequality

$$\mathbb{P}\left(\sup_{t\in[0,\delta]}\left|X_t^n - X_{t-}^n\right| > \eta\right) \geq \frac{k_e(A,B)}{m(A)}\mathbb{E}\left[\frac{\langle \tilde{Z}_0^n, eh_A\rangle}{1/(\delta n^2) + \langle \tilde{Z}_0^n, e\rangle + \langle \tilde{Z}_0^n, b\rangle/n + \langle \tilde{Z}_0^n, d\rangle/n}\right].$$

Under Assumption 4, one has that $\left(\langle \tilde{Z}_0^n, eh_A\rangle, \langle \tilde{Z}_0^n, e\rangle, \langle \tilde{Z}_0^n, d\rangle, \langle \tilde{Z}_0^n, b\rangle\right)$ converges to the deterministic limit $(\langle \xi_0, eh_A\rangle, \langle \xi_0, e\rangle, \langle \xi_0, d\rangle, \langle \xi_0, b\rangle)$ in probability thus one can get the a.s. convergence for a subsequence $\phi(n)$.
Moreover, since $\frac{\langle \tilde{Z}_0^n, eh_A\rangle}{1/(\delta n^2)+\langle \tilde{Z}_0^n, e\rangle+\langle \tilde{Z}_0^n, b\rangle/n+\langle \tilde{Z}_0^n, d\rangle/n} \leq 1$, the dominated convergence theorem leads to

$$
\begin{aligned}
&\limsup_{n\to+\infty} \mathbb{P}\left(\sup_{t\in[0,\delta]} \left|X_t^n - X_{t-}^n\right| > \eta\right) \\
&\geq \frac{k_e(A,B)}{m(A)} \lim_{n\to+\infty} \mathbb{E}\left[\frac{\langle \tilde{Z}_0^{\phi(n)}, eh_A\rangle}{1/(\delta\phi(n)^2) + \langle \tilde{Z}_0^{\phi(n)}, e\rangle + \langle \tilde{Z}_0^{\phi(n)}, b\rangle/\phi(n) + \langle \tilde{Z}_0^{\phi(n)}, d\rangle/\phi(n)}\right] \\
&= \frac{k_e(A,B)}{m(A)} \frac{\langle \xi_0, eh_A\rangle}{\langle \xi_0, e\rangle}.
\end{aligned}
$$

The choice $\epsilon = \frac{1}{2}\frac{k_e(A,B)}{m(A)}\frac{\langle \xi_0, eh_A\rangle}{\langle \xi_0, e\rangle}$, which is positive by Assumption 5 (iii), concludes the proof. $\diamond$

4 Examples and numerical illustration

4.1 Examples

We propose here to compute explicit solutions to Equation (2) for the stable composition. We analyze two examples: the two-dimensional mixture and the swap to the nearest neighbor. We omit here the dependence in age for clarity.

Mixture kernel. We want to derive explicit solutions to (2) in the case where $k_e(y,x)$ is a two dimensional mixture, that is $k_e(y,x) = b_1(y)f_1(x) + b_2(y)f_2(x)$ where b_1 and b_2 are continuous functions from $\mathcal{X}$ to $(0,1)$ such that $b_1 + b_2 \equiv \mathbf{1}$, and f_1 and f_2 are probability densities on $\mathcal{X}$. In this model, new characteristics are chosen based on two probability densities which are the same for all individuals but weights depend on the old characteristics of the individual. This is a particular case of a separable kernel with $n = 2$, and we illustrate the methodology described in Zemyan (2012) for the computations. We first derive the characteristic polynomial in $\lambda \in \mathbb{R}$ of the matrix

$$
A = \begin{pmatrix} \int b_1 f_1 & \int b_1 f_2 \\ \int b_2 f_1 & \int b_2 f_2 \end{pmatrix} = \begin{pmatrix} d_1 & 1-d_2 \\ 1-d_1 & d_2 \end{pmatrix},
$$

where $d_1 = \int b_1 f_1$ and $d_2 = \int b_2 f_2$. Then

$$
\begin{aligned}
P(\lambda) := |I - \lambda A| &= 1 - \lambda d_2 - \lambda d_1 - \lambda^2 + \lambda^2 d_2 + \lambda^2 d_1, \\
&= (1-\lambda)(1 + \lambda - \lambda d_2 - \lambda d_1).
\end{aligned}
$$

Since $P(1) = 0$, there exists non-trivial solutions to $(I - A)c = 0$. Let $Q(\lambda) = P(\lambda)/(1 - \lambda)$. Then $Q(1) = 2 - d_1 - d_2 > 0$ since $d_1 = \int f_1 b_1 < \int f_1 = 1$ (we assumed that $b_1, b_2 : \mathcal{X} \to (0, 1)$). We deduct that the space of all solutions is of dimension $p = 1$. Solutions of $(I - A)c = 0$ are of the form $c = \beta(1 - d_2, 1 - d_1)^T$, $\beta \in \mathbb{R}$. So that solution g to (2) is

$$g(x) = \frac{1}{\beta e(x)} \left((1 - d_2) f_1(x) + (1 - d_1) f_2(x) \right), \tag{13}$$

where $\beta = (1 - d_2) \int_{\mathcal{X}} \frac{f_1(x)}{e(x)} m(\mathrm{d}x) + (1 - d_1) \int_{\mathcal{X}} \frac{f_2(x)}{e(x)} m(\mathrm{d}x)$.

Swap to the nearest neighbor. We consider n classes, $\mathcal{X} = \{x_1, ..., x_n\}$. If a swap occurs for individual i with characteristic $x_i \in \{2, ..., n-1\}$, we suppose that its new one is uniformly chosen between x_{i-1} and x_{i+1}, that is $k_e(x_i, x')l(\mathrm{d}x') = \frac{1}{2}\delta_{x_{i-1}}(\mathrm{d}x') + \frac{1}{2}\delta_{x_{i+1}}(\mathrm{d}x')$. If the characteristics are $x_i = x_1$ or $x_i = x_n$, the neighbor is chosen with probability one, that is $k_e(x_1, x')l(\mathrm{d}x') = \delta_{x_2}(\mathrm{d}x')$ and $k_e(x_n, x')l(\mathrm{d}x') = \delta_{x_{n-1}}(\mathrm{d}x')$. The discrete formulation of Equation (2) is given by

$$\forall i \in \{1, ..., n\}, e(x_i)g(x_i) = \sum_{j \neq i} g(x_j)e(x_j)k_e(x_j, x_i).$$

Define the matrix A by $A_{i,j} = e(x_j)k_e(x_j, x_i)\mathbf{1}_{i \neq j} - e(x_i)\mathbf{1}_{i=j}$. We then want to find the eigenvectors for A associated with eigenvalue 0. Since characteristics changes can only lead to neighbor characteristics, the matrix A is tridiagonal. Remark also that upper and lower diagonals of A have same sign, which shows that it is diagonalizable. Let us write $A = PDP^{-1}$ the canonic decomposition with D diagonal. It is then straightforward to show that 0 is an eigenvalue for A (with multiplicity 1) and that an eigenvector Y is given by $Y_1 = \prod_{i \neq 1} e(x_i)$, $Y_n = \prod_{i \neq n} e(x_i)$ and if $k \in \{2, ..., n-1\}$, $Y_k = 2\prod_{i \neq k} e(x_i)$. So the solution g is given by

$$g(x_i) = \frac{Y_i}{\sum_{1 \leq k \leq n} Y_k}. \tag{14}$$

4.2 Numerical illustration

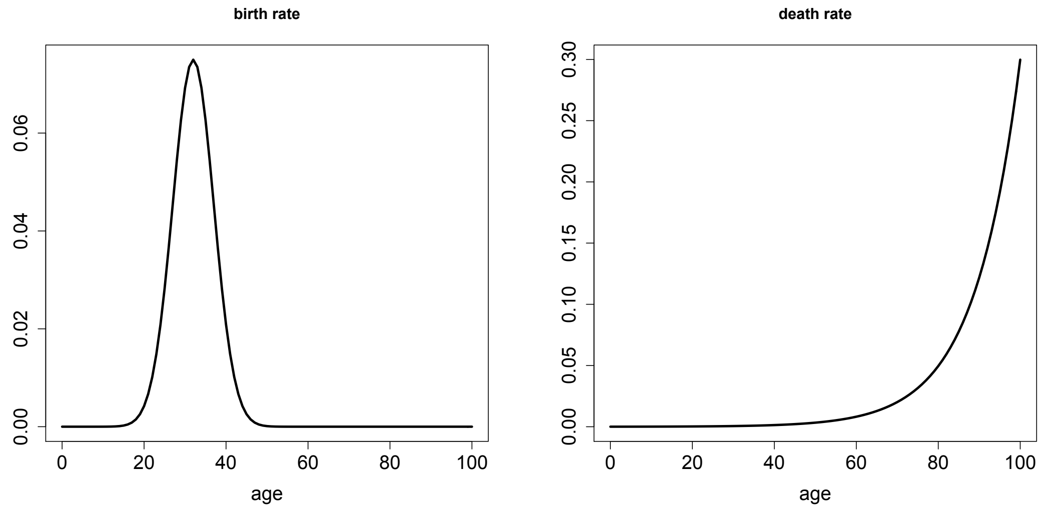
Numerical examples will be performed under the proportional hazard framework: $d(x, a) := \alpha(x)\bar{d}(a)$ and $b(x, a) := \beta(x)\bar{b}(a)$, with $\bar{d}(a)$ and $\bar{b}(a)$ some reference death and birth rate. In the macroscopic model, the death rate is

$$\hat{d}(a) = \int_{\mathcal{X}} d(x, a)g(x, a)m(\mathrm{d}x) = \hat{\alpha}(a)\bar{d}(a),$$

where $\hat{\alpha}(a) := \int_{\mathcal{X}} \alpha(x)g(x, a)m(\mathrm{d}x)$. In the same way, $\hat{b}(a) = \hat{\beta}(a)\bar{b}(a)$. It is interesting to note that in the general general framework where the invariant swap pattern is age-dependent, the equivalent model is not proportional hazard with an age-independent factor anymore.

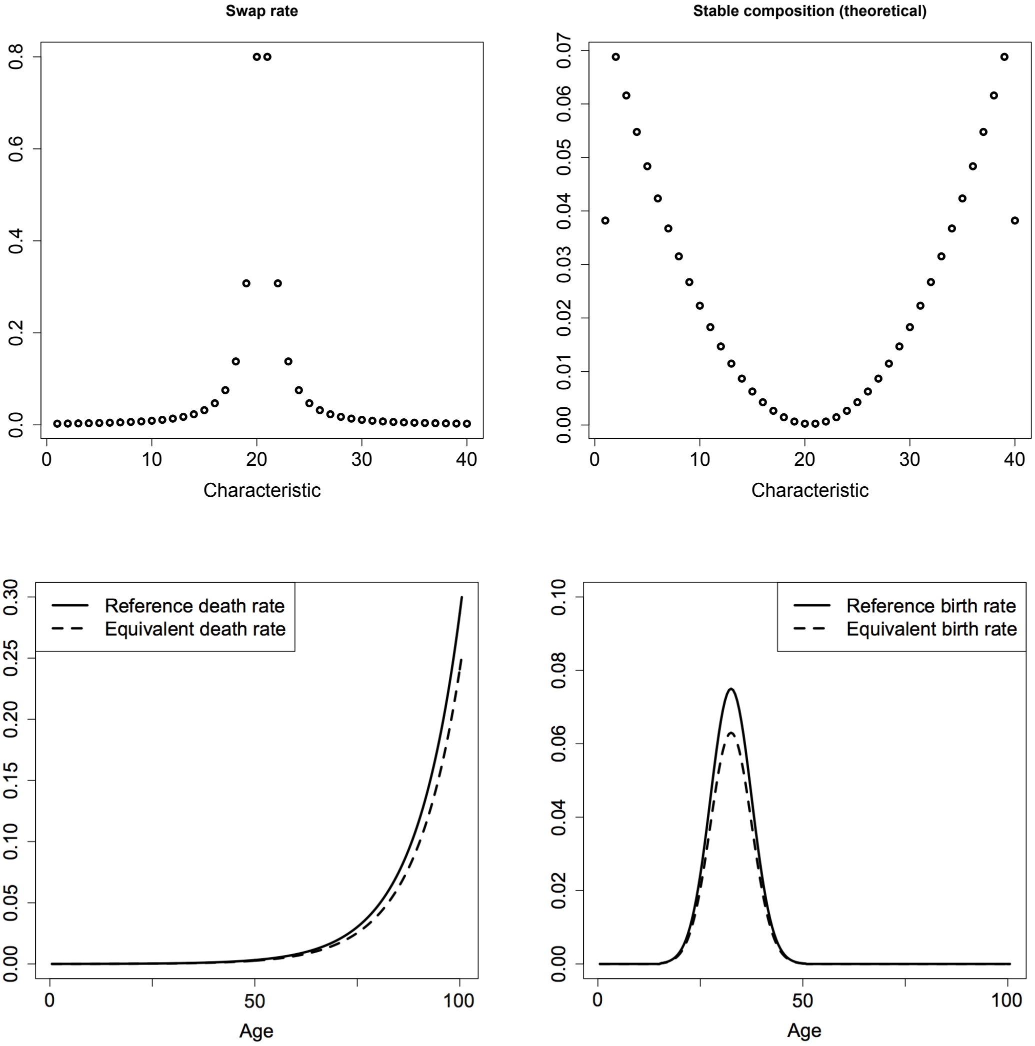
For the reference death rate, we choose a stylized Gompertz form $\bar{d}(a) = A_1 e^{B_1 a}$, and a stylized birth rate $\bar{b}(a) = A_2 e^{-B_2(a-\bar{a})^2}$. We choose arbitrary values that mimic the shape and the order of magnitude of human populations, inspired from death and birth data for France in 2008. For the birth data, the aim is to reproduce the shape of female birth rates and divide by two to approximately recover the intensity for an arbitrary individual (male or female). We take $A_1 = 3,7.10^{-5}$, $B_1 = 0.09$, $\bar{a} = 32$, $A_2 = 0.075$ and $B_2 = 0.02$. The reference rates $\bar{b}(a)$ and $\bar{d}(a)$ are plotted in Figure 1.

Figure 1: Stylized reference birth rate $\bar{b}(a)$ and death rate $\bar{d}(a)$ as a function of age.



We illustrate numerically the swap to the nearest neighbor. We consider $n = 40$ discrete classes and set $\mathcal{X} = \{1, ..., 40\}$. We choose the following dynamics : the caracteristic x is such that high value implies living longer but having less children, whereas a low value means having more children: let $\alpha(x) = \beta(x) = 1 - \frac{x-4,5}{100}$. The swap rate takes four possible forms, namely $e(x) = \frac{1}{1-(x-20,5)^2}$, $e(x) = \frac{1}{x}$, $e(x) = x/40$ and $e(x) = \sqrt{x/40}$. The swap rate, the theoretical stable composition and the equivalent birth and death rates are given for the four configurations from Figure 2 to 5. Let us emphasize that in the four experiments, the age and characteristic-dependent birth and death rates are the same, and also that the way the new characteristics are chosen at the time of swap is fixed (namely the swap to the nearest neighbor). We only vary the characteristics-dependent frequency of swap events, driving the way characteristics changes occur at the individual level. Let us first focus on the two upper graphs of each one of the four experiments. These show how the stable composition in terms of characteristics (right-upper graph) is linked to the swap pattern (left-upper graph). In each case, one can notice the interesting link between the shape of the swap rate and that of the stable composition given by the invariant measure. In particular, as expected, the characteristics for which the swap rate is high are less represented in the population composition, since individuals are forced to escape the corresponding class. Also, let us remark the side effect of the stable composition: this is due to the choice of the swap to the nearest neighbor mechanisms, since characteristics on the side are less likely to be chosen (they have one neighbor instead of two). Now, let us focus on the equivalent death and birth rates. First notice that these are always lower than the reference demographic rates, which are only used here to set a realistic age pattern; this is due to the fact that the proportional parameters $\alpha(x)$ and $\beta(x)$ are lower than one for almost all characteristics. One then gets insights on the impact of characteristics changes frequencies when comparing one by one the Figures 2 to 5. For example, the choice of the swap rate $e(x) = x/40$ leads to aggregate death and birth rates that are higher than in the three other configurations; this can be explained, as previously detailed, by the fact that small characteristics values, whose birth and death rates are higher, are over-represented. This shows how, with fixed swap rules and characteristic-specific demographic rates, the characteristic-dependent swap frequencies impacts what one observes at the macroscopic level.

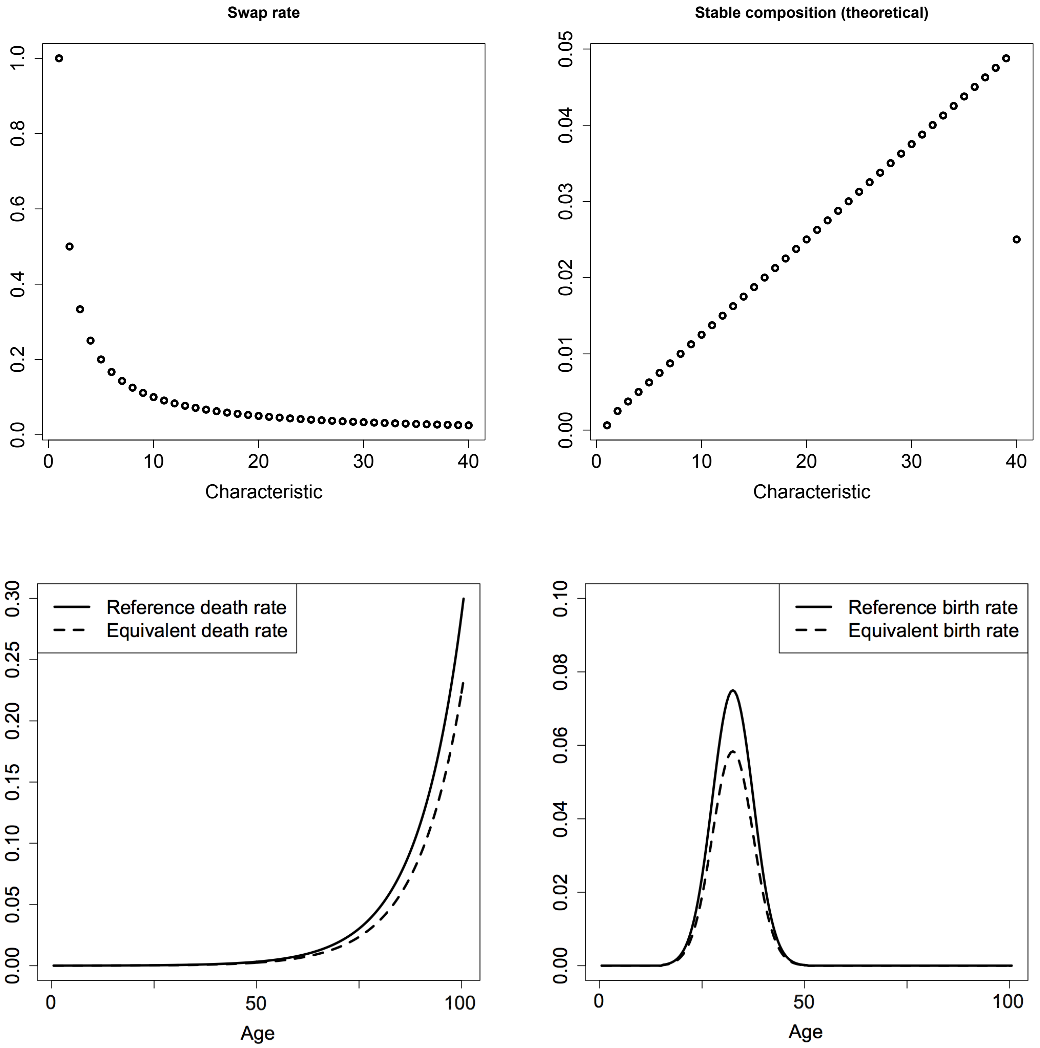
Figure 2: Swap rate, stable composition and equivalent death and birth rates for $e(x) = \frac{1}{1-(x-20{,}5)^2}$



5 Conclusion

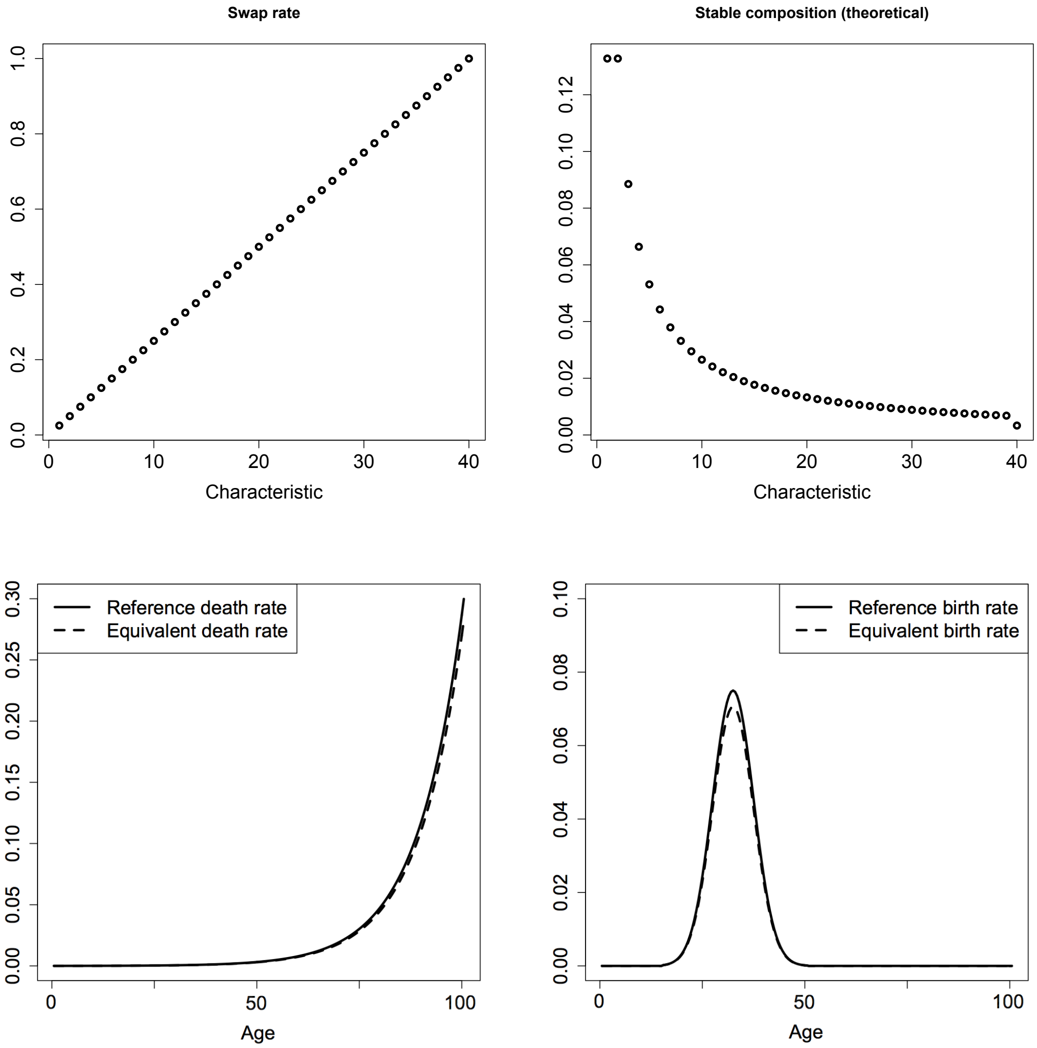
In this paper, we considered the large population limit of an age and characteristics-structured population model evolving according to individual birth, death and fast characteristic changes during life. Both the large population framework and the frequent characteristics changes appear naturally when focusing on the demographic evolution of a human population at the scale of a given country. When rescaling the population process, the classical deterministic transport-renewal McKendrick-Von Foerster equation appears, that describes the time evolution of the age pyramid. This dynamics is driven by age-dependent birth and death rates which are the average of microscopic birth and death rates over the stable characteristics distribution of the invariant swap pattern. In addition, we proved that it is not possible to keep

Figure 3: Swap rate, stable composition and equivalent death and birth rates for $e(x) = \frac{1}{x}$



track of the population structure in terms of characteristics, even in the case of age-independent demographic rates, in other words that the corresponding sequence of measure-valued processes is not tight. This gives a natural and interesting example of a sequence of processes which does not converge. A set of computable invariant distributions for the swap patterns have also been given, namely a simple mixture kernel and a swap to the nearest neighbor mechanism. When considering reference age-dependent birth and death rates that reproduce real demographic data, numerical illustrations have been performed showing the equivalent birth and death rates in a proportional hazard setting. Next steps in this direction could concern (i) the study of the associated central limit theorem, and the link with the speed at which the swap pattern reaches its invariant measure and (ii) the assessment of the numerical gain when considering the equivalent birth and death rates compared to the original model. These two aspects are linked in the assessment of the speed at witch

Figure 4: Swap rate, stable composition and equivalent death and birth rates for $e(x) = x/40$

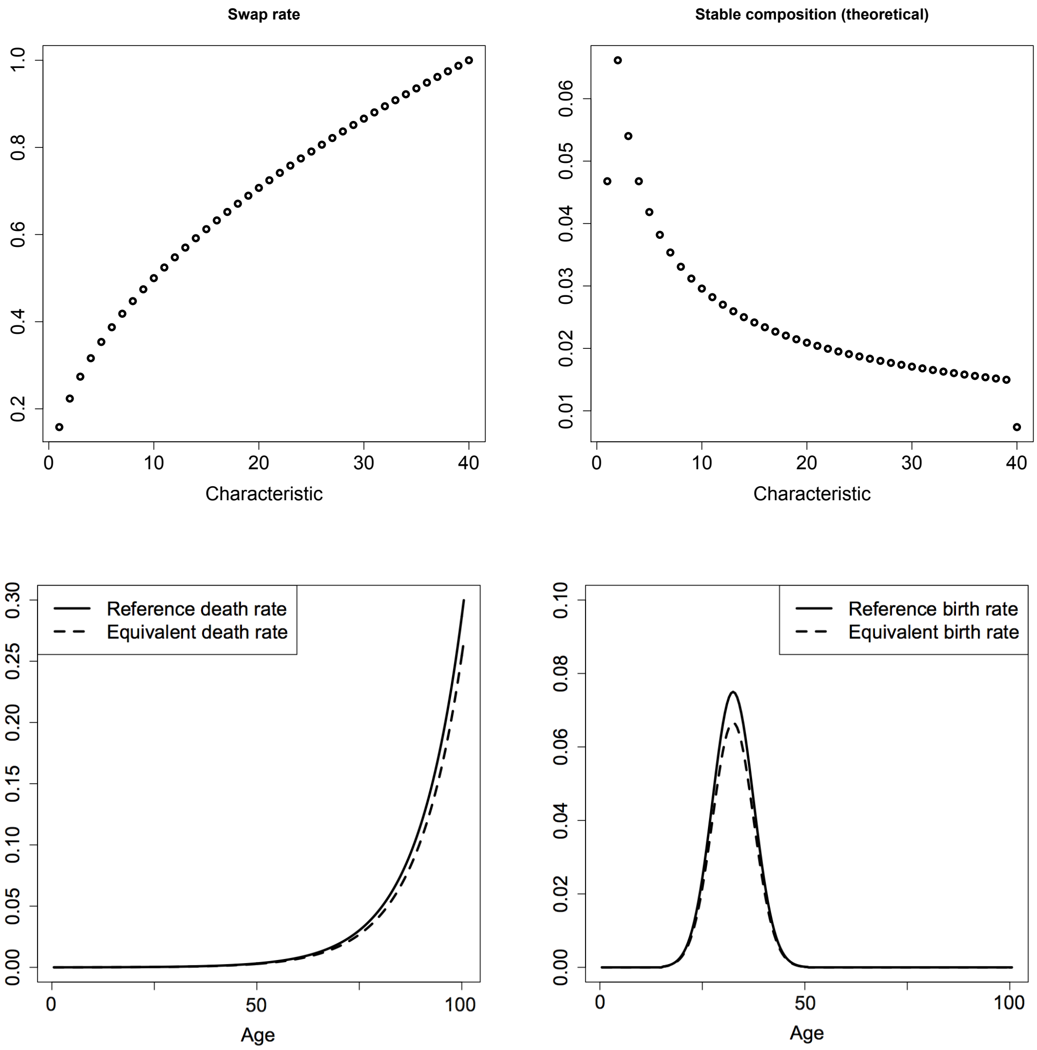


the original model reaches the averaged one. They are left for further research.

Acknowledgements

The author thanks his supervisors Nicole El Karoui and Stéphane Loisel for enlightening discussions, as well as fruitful discussions on particular topics with Mathieu Rosenbaum and Nicolas Gilliers on non-tightness issues, Thibault Bourgeron on steady states for partial differential equations, and also Pierre-Antoine Corre for his help and Casimir Emako for several interesting remarks.

Figure 5: Swap rate, stable composition and equivalent death and birth rates for $e(x) = \sqrt{x/40}$



References

Aldous, David. 1978. Stopping times and tightness. *The Annals of Probability* 335–340. 12

Auger, Pierre, JC Poggiale, E Sánchez. 2012. A review on spatial aggregation methods involving several time scales. *Ecological Complexity* **10** 12–25. 3

Bensusan, H. 2010. Risques de taux et de longévité: Modélisation dynamique et applications aux produits dérivés et à l'assurance vie. Ph.D. thesis, École Polytechnique. 3

Bensusan, H., A. Boumezoued, N. El Karoui, S. Loisel. 2010–2015. Bridging the gap from microsimulation practice to population models: a survey. *Work in progress* . 3, 4, 10

Billiard, Sylvain, Régis Ferrière, Sylvie Méléard, Viet Chi Tran. 2014. Stochastic dynamics of adaptive trait and neutral marker driven by eco-evolutionary feedbacks. *Journal* *of Mathematical Biology* 1–32. 3

Billingsley, Patrick. 2009. *Convergence of probability measures*, vol. 493. Wiley-Interscience. 14, 16, 20

Boumezoued, A., N. El Karoui, S. Loisel. 2015. Measuring mortality heterogeneity with multi-state models and interval-censored data. *HAL preprint* URL `https://hal.archives-ouvertes.fr/hal-01215350`. 2

Bovier, Anton, Shi-Dong Wang. 2013. Trait substitution trees on two time scales analysis. *arXiv preprint arXiv:1304.4640* . 3

Champagnat, Nicolas, Régis Ferrière, Sylvie Méléard. 2006. Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. *Theoretical population biology* **69**(3) 297–321. 2, 3, 4

Costa, Manon. 2015. A piecewise deterministic model for prey-predator communities. *arXiv preprint arXiv:1503.02492* . 3

Costa, Manon, Céline Hauzy, Nicolas Loeuille, Sylvie Méléard. 2015. Stochastic eco-evolutionary model of a prey-predator community. *Journal of mathematical biology* 1–50. 3

Ferriere, R., V.C. Tran. 2009. Stochastic and deterministic models for age-structured populations with genetically variable traits. *ESAIM: Proceedings*, vol. 27. 289–310. 2, 4

Fournier, N., S. Méléard. 2004. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Annals of applied probability* **14**(4) 1880–1919. 2, 3, 4

Gupta, Ankit, JAJ Metz, Viet Chi Tran. 2014. A new proof for the convergence of an individual based model to the trait substitution sequence. *Acta Applicandae Mathematicae* **131**(1) 1–27. 1, 3, 16

Joffe, Anatole, Michel Métivier. 1986. Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Advances in Applied Probability* 20–65. 12

Jourdain, Benjamin, Sylvie Méléard, Wojbor A Woyczynski. 2012. Lévy flights in evolutionary ecology. *Journal of mathematical biology* **65**(4) 677–707. 14, 18

Kurtz, Thomas G. 1992. Averaging for martingale problems and stochastic approximation. *Applied Stochastic Analysis*. Springer, 186–209. 9, 15, 16

McKendrick, A.G. 1926. Application of mathematics to medical problems. *Proc. Edin. Math. Soc.* **54** 98–130. 2

Méléard, Sylvie, Sylvie Roelly. 1993. Sur les convergences étroite ou vague de processus à valeurs mesures. *Comptes rendus de l'Académie des sciences. Série 1, Mathématique* **317**(8) 785–788. 13

Méléard, Sylvie, Viet Chi Tran. 2012. Slow and fast scales for superprocess limits of age-structured populations. *Stochastic Processes and their Applications* **122**(1) 250–276. 1, 3, 8, 9, 11, 14, 16

Roelly-Coppoletta, Sylvie. 1986. A criterion of convergence of measure-valued processes: application to measure branching processes. *Stochastics: An International Journal of Probability and Stochastic Processes* **17**(1-2) 43–65. 12

Tran, V.C. 2006. Modèles particulaires stochastiques pour des problèmes d'évolution adaptative et pour l'approximation de solutions statistiques . 7, 12

Tran, V.C. 2008. Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM: Probability and Statistics* **12**(1) 345–386. 2, 4, 10

Von Foerster, H. 1959. *The Kinetics of Cellular Proliferation*. Grune & Stratton. 2

Zemyan, Stephen M. 2012. *The Classical Theory of Integral Equations*. Springer. 6, 22